

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 2001, 21:53:05 : Search time 117 Seconds
(without alignments)
3289,483 Million cell updates/sec

Title: US-09-284-320-56

Perfect score: 1 gattccgagcgctgcaccc.....gttaatgatgatcccaaac 2033

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents, NA: *
1: /cgn1_7/ptodata/1/ina/5A.COMB.seq:*
2: /cgn1_7/ptodata/1/ina/5B.COMB.seq:*
3: /cgn1_7/ptodata/1/ina/6A.COMB.seq:*
4: /cgn1_7/ptodata/1/ina/6B.COMB.seq:*
5: /cgn1_7/ptodata/1/ina/PCNUS.COMB.seq:*
6: /cgn1_7/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	322	15.8	343	4	US-08-905-223-169
2	49.2	2.4	2781	3	US-08-749-522-4
3	47.4	2.3	6152	4	US-08-973-462-1
4	45.6	2.2	19124	2	US-08-487-826B-13
5	43.8	2.2	5852	1	US-07-867-106-2
6	43.8	2.1	5852	1	US-07-867-106-2
7	43.4	2.1	827	4	US-08-998-416-535
8	43	2.1	1618	1	US-07-885-970A-24
9	43	2.1	1618	1	US-08-298-687A-24
10	43	2.1	1618	1	US-08-530-797-14
11	43	2.1	1618	1	US-08-998-829-24
12	43	2.1	1618	1	US-08-787-335-14
13	42.6	2.1	3701	4	US-08-845-258-10
14	42.6	2.1	3701	4	US-08-990-571-10
15	42	2.1	731	1	US-08-451-405A-2
16	41.2	2.0	1507	3	US-08-605-150A-15
17	40.8	2.0	5176	4	US-08-654-482-13
18	40.6	2.0	1395	1	US-07-991-867B-25
19	40.6	2.0	1395	1	US-08-107-755A-25
20	40.6	2.0	1395	2	US-08-544-332-25
21	40.6	2.0	6768	1	US-08-107-755A-1
22	40.6	2.0	8457	1	US-07-991-867B-1
23	40.6	2.0	8457	2	US-08-544-332-1
24	40.4	2.0	636	4	US-08-998-416-1137
25	40.2	2.0	740	4	US-08-998-416-563
26	40	2.0	615	4	US-08-998-416-186
27	40	2.0	19124	2	US-08-487-826B-13

C	28	39.6	1.9	1422	1	US-08-319-704-5	Sequence 5, Appl1
C	29	39.4	1.9	837	4	US-08-998-416-288	Sequence 288, App
C	30	39.2	1.9	665	2	US-08-883-795A-36	Sequence 36, Appl
C	31	38.8	1.9	10342	4	US-08-972-927-5	Sequence 971, Appl
C	32	38.4	1.9	740	4	US-08-998-416-971	Sequence 9, Appl
C	33	38.4	1.9	3701	4	US-08-845-258-10	Sequence 10, Appl
C	34	38.4	1.9	3701	4	US-08-990-571-10	Sequence 1, Appl1
C	35	38.4	1.9	6243	2	US-09-056-075-1	Sequence 7, Appl1
C	36	38.4	1.9	8868	3	US-08-815-809-7	Sequence 8, Appl1
C	37	38.4	1.9	19877	2	US-08-816-155B-8	Sequence 1137, Ap
C	38	38.4	1.9	19877	3	US-09-079-587-8	Sequence 1, Appl1
C	39	38.2	1.9	636	4	US-08-998-416-1137	Sequence 5, Appl1
C	40	38.2	1.9	26700	1	US-08-472-217-1	Sequence 14, Appl
C	41	38.2	1.9	26700	2	US-08-488-199-5	Sequence 1, Appl1
C	42	38.2	1.9	26700	3	US-08-760-534A-1	Sequence 14, Appl
C	43	38	1.9	7218	1	US-08-232-463-14	Sequence 2, Appl1
C	44	37.8	1.9	2210	1	US-07-710-361-2	Sequence 5, Appl1
C	45	37.8	1.9	2317	3	US-08-749-522-5	

ALIGNMENTS

RESULT 1
US-08-905-223-169
Sequence 169, Application US/08905223
Patent No. 6222029
GENERAL INFORMATION:
APPLICANT: Edwards, Jean-Baptiste D.
APPLICANT: Duclert, Aymeric
APPLICANT: Lacroix, Bruno
TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
NUMBER OF SEQUENCES: 503
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 169:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE: Homo Sapiens
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: other
LOCATION: 104..336
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: Identity 96
OTHER INFORMATION: region 1..233
OTHER INFORMATION: id H07998

APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/A990/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match 2.2%; Score 43.8; DB 1; Length 5852;
Best Local Similarity 49.8%; Pred. No. 0.055;

Matches 111; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 1202 aaaaatacctttagtgcgttaagtagatgaatcattcaataaataa 1261
DB 2157 AATATTATACATTTTGGCAGACATTTTATTTATTTATTTTGAATTTT 2216
QY 1262 tcaatttgccttatttgcgtgcgcgtgaatcttctctagaatga 1321
DB 2217 TTTTATTTTAAATTTTCTTTTCTTTTCTTTTAAATTTTAAATTTT 2276
QY 1322 ttgacgtgaatccacgtgtagatccataataatgctgaatataatagcc 1381
DB 2277 TTTCCACACATTTTATTTTATTTTATTTATTTGTAATTCATTTTATTTT 2336
QY 1382 atttaatacatgattcattctctgttaagaattggaaat 1424
DB 2337 AATTAATAGTTTGGTTAATTTATTTCAAGATTTTAAAAAT 2379

RESULT 6

US-07-867-106-2/c
Sequence 2, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Shade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526r1s
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/A990/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match 2.2%; Score 43.8; DB 1; Length 5852;
Best Local Similarity 44.8%; Pred. No. 0.055;

Matches 168; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

QY 1167 attagaagaaggggcttggaattgctgttctgtaaaatatactttagtgcctta 1226
DB 2150 AATATCAATTTTATTAATTAAGTAATTAATTAATTAATTAATTAATTAATTA 2091
QY 1227 aatgagatgatacttcaattataaaataaataaatttgccttatttgcgt 1286
DB 2090 AAAAAAATCAAAAAAACAAGTAATTTATTTATTAAGAGGCTTTTATTTT 2031
QY 1287 gtcctgtgaatcttctctagaggaattatagatgacgtgaatccacgtgata 1346
DB 2030 TTTTATTTTCTTTTCTTTTCAAGTAAAAAATAAAAAAATAAAAAAATAAAAA 1971
QY 1347 gatccataatgctgaatattatgatatagccatttaacattgattcattcg 1406
DB 1970 AGTTGGTTAACTACATTTATTTATTTATTTTATTTTGAATTTTAAATTTAATTT 1911
QY 1407 ttaatagaattggaatatgacacgaagaatgtaaaacatttagaataagcgtgt 1466
DB 1910 TTAATATTTTATTTATTAATTAAGATCTAATTAATAAAAAAATAAAAAAATAAAAA 1851
QY 1467 atgaaaaaagtcacgaattattatagacaacttaagaaatgaatgacttcttaca 1526
DB 1850 AAGAAAAAATAAAAAAAGTAGAATTTATTTAAATTTAAATTTATTTCAATCTTAATA 1791
QY 1527 gcatggtgaatc 1541
DB 1790 AATTAAGTATATATC 1776

RESULT 7

US-08-998-416-535
Sequence 535, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine

Oy	1458	gcgcgctcctatggaaaaagcgcctgaatttatggacaactgaagcttacct	1511	
Db	576	tcttttttaattatattttaaattcattttatataagaaattatattattattttattatcaat	635	
Oy	1518	ctttacacagcatalgltgaaacatcatattggcgtcatgtcact	1562	
Db	636	aatttttgattaatatattaccattattattatggnntttatanaat	680	

GENERAL INFORMATION:
APPLICANT: John, Maliyakal E.
TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
NUMBER OF SEQUENCES: 33

ADDRESSEE: Nicholas J. Seay, Quarles & Brady
STREET: P.O. Box 2113, First Wisconsin Plaza
CITY: Madison

COUNTRY: USA
ZIP: 53701

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/885,970A
 ; FILING DATE: 19920518
 ;

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/617,239
FILING DATE: 21-NOV-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/253,243
FILING DATE: 04-OCT-1988

NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
TELECOMMUNICATION INFORMATION:

TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1618 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
;

```

MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO

ORGANISM: *Cebia pentandra*
IMMEDIATE SOURCE:

CLONE: E6 K3
US-07-885-970A-24

Query Match	2.1%;	Score
Best Local Similarity	55.8%;	pred.
Matches 82: Conservative	0;	Mis

QY 1180 gttggaattgctgttltgttaaatatata
| | | | | | | | | | | | | | | | | |

QY 1240 accttaccattataaaaaaatcaaattt
| | | | | | | | | | | | | | | |

Db 89 CCATTTT TTTATATAAATAATCCTAATCT

QY 1300 ttcttagagtgaaattatagatgac 1326
| | | | | | | | | | | | | | | | | |
Db 29 TATTAAATTTAAATATCTAATATC 3

RESULT 9

US-08-298-687A-24/C
; Sequence 24, Application US/08298687A
; Patent No. 5521078
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; APPLICANT: Umbeck, Paul F.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
; TITLE OF INVENTION: FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,687A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1618 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Cebia pentandra
; IMMEDIATE SOURCE:
; LIBRARY: EMBL-CP
; CLONE: E6-R5
; US-08-298-687A-24

Query Match 2.1%; Score 43; DB 1; Length 1618;
Best Local Similarity 55.8%; Pred. No. 0.045;
Matches 82; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1180 gtggaaatgctgttggttaaaatatacttttagtggtcctttaagtagatgac 1239
| | | | | | | | | | | | | | | | | |
Db 149 GATGACCTTATGATTTCTTCTTTATTTATTTATTTGTCCTTTTATGATCTCATTT 90
QY 1240 accttaacattataaaaaaatacaatttgcttcttatttggtgctgctgtagt 1299
| | | | | | | | | | | | | | | | | |
Db 89 CCATTTTATTAATAAATAAATCAATCAATTTTATTTTATTTTATTTTATTTTAA 30
QY 1300 ttcttagagtgaaattatagatgac 1326

Db 29 TATTAAATTTAAATATCTAATATC 3
| | | | | | | | | | | | | | | | | |

RESULT 10

US-08-530-797-14/C
; Sequence 14, Application US/08530797
; Patent No. 5597718
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; APPLICANT: Umbeck, Paul F.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
; TITLE OF INVENTION: FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O. Box 2113
; CITY: MADISON
; STATE: WISCONSIN
; COUNTRY: U.S.A.
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,797
; FILING DATE: 20-SEP-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/617,239
; FILING DATE: 21-NOV-90
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-88
; ATTORNEY/AGENT INFORMATION:
; NAME: Nicholas J. Seay
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990245
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1618 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHEICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Cebia pentandra
; STRAIN: kapok
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: EMBL-CP
; CLONE: PSKCP6-RV
; US-08-530-797-14

Query Match 2.1%; Score 43; DB 1; Length 1618;
Best Local Similarity 55.8%; Pred. No. 0.045;
Matches 82; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1180 gtggaaatgctgttggttaaaatatacttttagtggtcctttaagtagatgac 1239
| | | | | | | | | | | | | | | | | |
Db 149 GATGACCTTATGATTTCTTCTTTATTTATTTATTTGTCCTTTTATGATCTCATTT 90
QY 1240 accttaacattataaaaaaatacaatttgcttcttatttggtgctgctgtagt 1299
| | | | | | | | | | | | | | | | | |
Db 89 CCATTTTATTAATAAATAAATCAATCAATTTTATTTTATTTTATTTTATTTTAA 30
QY 1300 ttcttagagtgaaattatagatgac 1326

Db 29 TATTAAATTTAAATACTAAATAC 3

RESULT 11

US-08-298-829-24/c
Sequence 24, Application US/08298829
Patent No. 5620882
GENERAL INFORMATION:
APPLICANT: John, Mallyakal E.
TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nicholas J. Seay, Charles & Brady
STREET: P.O. Box 2113, First Wisconsin Plaza
CITY: Madison
STATE: Wisconsin
COUNTRY: USA
ZIP: 53701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/298, 829
FILING DATE: 19-OCT-1994
CLASSIFICATION: 800
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/885, 970
FILING DATE: 18-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/617, 239
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/253, 243
FILING DATE: 04-OCT-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 283-2478
TELEFAX: (608) 251-5139
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1618 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Cebia pentandra
IMMEDIATE SOURCE:
LIBRARY: EMBL-CP
CLONE: E6-R5
US-08-298-829-24

Query Match 2.1%; Score 43; DB 1; Length 1618;

Best Local Similarity 55.8%; Pred. No. 0.045; Mismatches 65; Indels 0; Gaps 0;

Db 1180 gtggaaattggcgtgttggtaaatatatacttttagtggctttaaagtagatgat 1239
149 GATTGACTTTGATTTCTTCTTATATATATATATTTGGTCTCTTTTATGATCTCATTT 90
Qy 1240 acttaacattataaaaaaaatcaaatcttgcttatttggtgctgctggaagt 1299
89 CCAATTTTTTATAAATAAATCTAATCTAATTTTTTTTTTAATTTTACATATATTA 30

Qy 1300 ttctcagagtgaaatcatagatgac 1326
Db 29 TATTAAATTTAAATACTAAATAC 3

RESULT 12

US-08-787-335-14/c
Sequence 14, Application US/08787335
Patent No. 5981834
GENERAL INFORMATION:
APPLICANT: John, Mallyakal E.
APPLICANT: Umbeck, Paul F.
APPLICANT: Brill, Winston J.
TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
TITLE OF INVENTION: FOR ALTERED FIBER
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles and Brady
STREET: P.O. Box 2113
CITY: MADISON
STATE: WISCONSIN
COUNTRY: U.S.A.
ZIP: 53701
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/787, 335
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/530, 797
FILING DATE:
APPLICATION NUMBER: US 07/253, 243
FILING DATE: 04-OCT-88
ATTORNEY/AGENT INFORMATION:
NAME: Nicholas J. Seay
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 1122990245
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1618 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Cebia pentandra
STRAIN: Kapok
TISSUE TYPE: fiber cells
IMMEDIATE SOURCE:
LIBRARY: EMBL-CP
CLONE: PSKCP6-RV
US-08-787-335-14

Query Match 2.1%; Score 43; DB 2; Length 1618;

Best Local Similarity 55.8%; Pred. No. 0.045; Mismatches 65; Indels 0; Gaps 0;

Db 1180 gtggaaattggcgtgttggtaaatatatacttttagtggctttaaagtagatgat 1239
149 GATTGACTTTGATTTCTTCTTATATATATATATTTGGTCTCTTTTATGATCTCATTT 90
Qy 1240 acttaacattataaaaaaaatcaaatcttgcttatttggtgctgctggaagt 1299
89 CCAATTTTTTATAAATAAATCTAATCTAATTTTTTTTTTAATTTTACATATATTA 30

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 2001, 22:01:01 ; Search time 5436.15 Seconds
(without alignments)
5784.596 Million cell updates/sec

Title: US-09-284-320-56

Sequence: 1 gagtcgagcgcgctcacctc.....gttaatgatgatcccaaac 2033

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pi1:*
13: gb_pi2:*
14: gb_pi3:*
15: gb_pi4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_higo_hum:*
20: em_higo_inv:*
21: em_higo_rod:*
22: em_hig_hum1:*
23: em_hig_hum2:*
24: em_hig_hum3:*
25: em_hig_hum4:*
26: em_hig_hum5:*
27: em_hig_hum6:*
28: em_hig_hum7:*
29: em_hig_hum8:*
30: em_hig_inv1:*
31: em_hig_inv2:*
32: em_hig_other:*
33: em_hig_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pi1:*
48: em_pi2:*
49: em_pi3:*
50: em_pi4:*
51: em_pi5:*
52: em_pi6:*
53: em_pi7:*
54: em_pi8:*
55: em_pi9:*
56: em_pi10:*
57: em_pi11:*
58: em_pi12:*
59: em_pi13:*
60: em_pi14:*
61: em_pi15:*
62: em_pi16:*
63: em_pi17:*
64: em_pi18:*
65: em_pi19:*
66: em_pi20:*
67: em_pi21:*
68: em_pi22:*
69: em_pi23:*
70: em_pi24:*
71: em_pi25:*
72: em_pi26:*
73: em_pi27:*
74: em_pi28:*
75: em_pi29:*
76: em_pi30:*
77: em_pi31:*
78: em_pi32:*
79: em_pi33:*
80: em_pi34:*
81: em_pi35:*
82: em_pi36:*
83: em_pi37:*
84: em_pi38:*
85: em_pi39:*
86: em_pi40:*
87: em_pi41:*
88: em_pi42:*
89: em_pi43:*
90: em_pi44:*
91: em_pi45:*
92: em_pi46:*
93: em_pi47:*
94: em_pi48:*
95: em_pi49:*
96: em_pi50:*
97: em_pi51:*
98: em_pi52:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	2028.8	99.8	2049	89	AF248966 Homo sapi
2	1874.8	92.2	1884	93	HSW000272 Homo sapi
3	1429.8	70.3	158239	80	AL356315 Homo sapi
4	1429.8	70.3	214692	81	AL513325 Homo sapi
5	1428.2	70.3	165617	80	AL356381 Homo sapi
6	1078	53.0	126543	69	AC026156 Homo sapi
7	622	30.6	622	97	HSY17975 Homo sapien
8	378.6	18.6	513	54	G27225 human SRS S

Db 912 TTGGACACCTCCCTCATTAGGAAGACAGCTATCTTGAGGCAAAACGACGACAGAAC 971
 Oy 961 ccagcaagtcctcctaacccttgacataagataatttgaatattccgtggttttcaac 1020
 Db 972 CCAGCAAGTCCCTTAATACCTTGATATAGATATATTTTGAATATTTCCGTGTTTCAAC 1031
 Oy 1021 atggtacttggaatgaatgacgcttgcttgcttgcttgcttgcttgcttgcttgcttgct 1080
 Db 1032 ATGGTACTTGGATTAATGATCGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCTTGCT 1091
 Oy 1081 tggacacatgacatcctgataatgataatcattatagaatgacacacagaatgcga 1140
 Db 1092 TGGACATGATGATCTGGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1151
 Oy 1141 atggtacttggaatgacgcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgct 1200
 Db 1152 ATGGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1211
 Oy 1201 taaatatacctcttgatgcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgct 1260
 Db 1212 TAAATATATCTTTAGTGTGCTTAAAGTATGATGATGATGATGATGATGATGATGATGATGAT 1271
 Oy 1261 atcaaatcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgct 1320
 Db 1272 ATCAAAATTTGCTTTATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1331
 Oy 1321 attgacgtgaatcccaactgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1380
 Db 1332 ATTGACGTGAATCCCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1391
 Oy 1381 cattataatgaatgatttattcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgct 1440
 Db 1392 CATTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1451
 Oy 1441 gtaaacattagaatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1500
 Db 1452 GTAAACATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1511
 Oy 1501 ttgacgaatgaatgatttattcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgct 1560
 Db 1512 TTGACGATGCTTAACTTCTTTACACGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 1571
 Oy 1561 ctatgaacaatttgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1620
 Db 1572 CTATGACATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1631
 Oy 1621 ttttttaactgaatgacgcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgct 1680
 Db 1632 TTTTAACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1691
 Oy 1681 cgtatcgtgagtaacagagagagagagagagagagagagagagagagagagagagagagagag 1740
 Db 1692 CTGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1751
 Oy 1741 acatggagcgaatgaatgataatgataatgataatgataatgataatgataatgataatgataat 1800
 Db 1752 ACATGGGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1811
 Oy 1801 gagtatataatgatttattcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgcttgct 1860
 Db 1812 GAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1871
 Oy 1861 aatataaacttatttatttatttatttatttatttatttatttatttatttatttatttatttattt 1920
 Db 1872 AATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1931
 Oy 1921 tacaactatctatagaaatagaatagaatagaatagaatagaatagaatagaatagaatagaataga 1980
 Db 1932 TACAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1991
 Oy 1981 gtggacacatgctatcaagagagagagagagagagagagagagagagagagagagagagagagag 2032
 Db 1992 GTGACCATGCTATCAAGAGTACAAATTAAGTTAATGATGATGATGATGATGATGATGATGATGAT 2043

RESULT 2
 HSM800272
 LOCUS
 DEFINITION
 Homo sapiens mRNA; cDNA DKFZp56400582 (from clone DKFZp56400582);
 partial cds.
 AL049929.1
 accession
 version
 keywords
 SOURCE
 ORGANISM
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 Ponsita, A., Klein, M., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Submitted (15-MAY-1999) MIPS, Am Klopferspitz 18a, D-62152
 Martinsried, GERMANY
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 Heidelberg/Germany) within the CDNA sequencing consortium of the
 German Genome Project.
 This clone (DKFZp56400582) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at <http://www.mips.biochem.mpg.de/proj/cdna/>.
 Location/Qualifiers
 FEATURES
 source
 1..1884
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /dev_stage="fetal"
 /tissue_type="brain"
 /clone_lib="DKFZp56400582"
 /clone_lib="564 (synonym: hfdp2). Vector pAMP1; host
 XL-2blue; sites NotI + SalI"
 /map="X"
 1..995
 /gene="DKFZp56400582"
 <1..995
 /gene="DKFZp56400582"
 /note="weak similarity to C.elegans T24D5.2"
 /codon_start=3
 /product="hypothetical protein"
 /protein_id="CAB43210.1"
 /db_xref="GI:4884175"
 /translation="TIRKPSGVFRNGMNPTRGPIRDPVAALSMGFSYKEDLSNPGLA
 VGNLFRPRARYVMVAVKVKYKALPGSIVSYRLENAPPSLDVANSISHSLSSEETP
 VVLDLASEERYVMGRANSVFEDLSYTLRQLRNLFQVENSISLPLNLSRNNED
 LFLISELYVLDHDISSLRSRHLAKDHSPLDLSLEGLDLEIKRRIGDSQFPDASK
 LLDVALDKPADMDYSILYGNAAVVELTVKSPFSLIRKPTTILEAARAANPASPYNIA
 YKYNFEYSVVFNMVLMIALALAVIITSYNIMNDPXDYSIIIRYMTNOKIRMP"
 polyA_signal
 1854..1859
 polyA_site
 1874
 BASE COUNT 555 a 322 c 393 g 614 t
 ORIGIN
 Query Match 92.2%; Score 1874.8; DB 93; Length 1884;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1876; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 155 gttatataatcacccagggtcgtgtgttttccgaatagaaatggcctaccagag 214
 Db 1 GTATATTAATAATCACAGGCTGTTGTTTCCGAATGGAATG6CCTATACAGAG 60
 Oy 215 agcgcataccagacgtgctgcatactgcatggctctctgtgaagaagaccttct 274
 Db 1 AGCGATCCCAAGACGCTGCTGATGTCATGGCTTCTGTGAAGAAGACCTTTCT 120
 Oy 275 ggcacgactgcagtggtgaactgttccatcgtctcctcggtaccgctacgtgctg 334

```

|||||
Db 121 GGCAGACGCTGGCACTGGTACCTGTTTCATCGCCGAGGCTACCGCTACGTCATGTCATG 180
Oy 335 tgaaggaaatgaaacaactggcctcctaccccaagcagtgcaatctcgtaecccttgaga 334
Db 181 TGAAGGAGAGTGAACAACTGGCTCTACCCCGAGGAGTCATTTGTTACCTCTTGGAGA 240
Oy 395 atgcaatcctcttgaactgagcaagtgatgcaaatcctcctcctccttattctcga 454
Db 241 ATGACAGTCTCTTTAGTCTTGACAGTGTGCAAAATTCATTCACCTCTTATTTCTGAGG 300
Oy 455 aaactcctgttcttgagatggcctcccaagtgagaaagatgatatgtaggaag 514
Db 301 AAATCCTCTGTGTGTGAGTGGCTCCAGTGAAGAAAGATGATATGAGGAGAG 360
Oy 515 caaactcaagtgttgaagaccttgcagtaactcgcagctcgcgtaacgctgttc 574
Db 361 CAAACCTGAGTGTGTAAGACCTTTCAGTCACTGGCGCAGCTCCGTAACTCCCTGTTTC 420
Oy 575 aagaagaactctgtctcaagtcaactccctcaattctcagtagaagaatgaagtg 634
Db 421 AAGAAACCTGTGTCTGAGTCACTCCCTCAATTCCTGAGTGAACCAATGAAGTTG 480
Oy 635 acctgctcttcttgaactgcaagtgcatactgatalcttaagctgtcgtcgtc 694
Db 481 ACCGCTCTTCTTCTGAACTGCAAGTGTCTACATGATTTTCACCTTGGCTCTGTC 540
Oy 695 atagatctatgcaagaagatcatctccgattcttatctacagtgagctgcaagtttg 754
Db 541 ATAGAGATCTTACCAAGATCATCTCTCGATTTATATACGAGGAGCTGGCAGGTTGG 600
Oy 755 atgaatctggaagagctatgaggaagacactgcaaatcagaagatgcttlaagatc 814
Db 601 ATGAATTTGGAGAGGTTATGAGGAGATCTGAAACAATTCAGAGATGCTTCTAAGTTC 660
Oy 815 ttgttaagctcttgaagaatttgagatgacatgtaactcttatagttggaatgcag 874
Db 661 TTGTATACCTCTGCAAAAGTTTGCAGATGACATGTACAGTCTTTATGGTGGAGATCAC 720
Oy 875 tggtagagttagcaactgcaatgcatcttgaacacctccctcattagaagaagaagact 934
Db 721 TGGTAGAGTTAGTCACTGTCAAGTCATTTGACACCTCCCTCATTTAGAGAGACATG 780
Oy 935 tcccttaggcaaaacagaagaagaaccagaagctcccttaactcctgataatagata 994
Db 781 TCCCTTAGGCAAAACGACGAGACACCCAGAGCCCTATTAACCTTGATATTAAGTAT 840
Oy 995 atttgaatattcgttggttttcaacatgtaacttggataatgacgcttgcttg 1054
Db 841 ATTTTGAATATTTCCGTGTTTCAACATGTGTAATTTGGATTAATGATCGCTTGCCCTGG 900
Oy 1055 ctgtgattatacctcttacaatatttgaagaatgagatcctgataatagataatcatt 1114
Db 901 CTGTGATTTATCACTCTTAAATATTTGGACATGAGATCTGTGATATGATAGCATCTT 960
Oy 1115 atagagatgcaaacagaagatcgaaatggaatggaatgtaaccgtgccaataga 1174
Db 961 ATGGATGCAAAACACAAATTCGAATGGAATGAATGATTAACCTGCGCAAAATTAAGAA 1020
Oy 1175 agggagttggaatctgctgttgtttaaataatatttgaatgcttgaatgaat 1234
Db 1021 AGGGGTTGGAAATTTGGCTGTTTGTGTAATAATATTTTGTAGTGTGCTTTAAGAT 1080
Oy 1235 agtaatacttcaatataaanaaaatcaaatgttgccttatttctgtgtgcctgt 1294
Db 1081 AGTATCTTTCATTTATTAATAAAATCAAAATTTGCTCTTATTTTGTGTGCTGCTG 1140
Oy 1295 gatgtttcttagagtaattatgattgacgtgaatcccaactggtatagaatccat 1354
Db 1141 GATGTTTTCAGAGTGAATTAATAGATTGACGTGAATCCACCTGTGTATAGATTTCAT 1200
Oy 1355 aataatgctgaatatatgatatagcaatlaataacatgatttcaattctgtttaa 1414
|||||

```

```

Db 1201 AATATGCTTGATATTAATGATATAGCATTTAATACATTTGATTCATTCTTTAATGC 1260
Oy 1415 atttgaanaalagcaactgaaagaalgtanaacattagaatagctogltatganaa 1474
Db 1261 ATTTGAAATATATGACATGAAAGAAATGTAATAACATTTAGATAGCTGTTATGAGAAA 1320
Oy 1475 aatgcaactgaattatatacaaaacttaagaaatgcaatgtaacttcttacaagaatagt 1534
Db 1321 AATGACATGAAATTTATTAACAAACTTACAAATTCCTTAACCTTTTACACAGATAGGT 1380
Oy 1535 gaaatcatatttgggctatltgatactatgaaacaatttgaatgtaatglttaattgagt 1594
Db 1381 GAAATACATATTTGGGCTATTGTATATCATGAAACAATTTGTAAATGCTTATTTATGTGT 1440
Oy 1595 aaataactctgaaacaagaagaagaattttaaacttgaagtagccctaaatatgagat 1654
Db 1441 AATATACCTGTAACACAGAAAGAAAGTTTAACTTAGACTAGCCCTAAATATGATGT 1500
Oy 1655 gcttataatagccttagtlttggaaactgtatctgaatacagaggacagctgttltta 1714
Db 1501 GCTTATATTAATTCGCTTGTGTTTGAACCTGTATCTGATGAACAGACAGCTGTTTTTTA 1560
Oy 1715 acctctctgcaagttgttgaacctacatgagtaataatgataactaaataactacat 1774
Db 1561 ACCCTTCTGCAAGTTTGTATACCTACATGCGCTAATATGATATGATTAATAATACAT 1620
Oy 1775 tgaatcaagaagaactagcctgtgtgagatatalagatgcttccattatacaacaaa 1834
Db 1621 TGATCTAAGAAAGAAACATGACCTTGTGAGATATATAGATGCTTTTCAATTAATACACAAA 1680
Oy 1835 atccctgaggaacatttggagcatgaaatacaaacatttattcagtaacttccc 1894
Db 1681 ATCCCTGAGGAGCATTTTGGAGCATGATATTAACATTTTATTTAGTACATCTTTCC 1740
Oy 1895 cctgtgaagttaactatggttgtgtgatacacttcaactctataagaatgaatggaat 1954
Db 1741 CCTGTATAGTACTATGATGTTGTGTGTCAACTTCATCTATGATTAAGATGAGAACT 1800
Oy 1955 ggtgaattcacttcttlatgttggagtgagcaatgctcatalcaagaagtgaacaataag 2014
Db 1801 GGGTGAATTTACTTTTATGTGGAGTGGACCAATGTCTATCAAGACTGACAAATTAAG 1860
Oy 2015 ttaatgatgattccaaa 2032
Db 1861 TTAATGATGATTTCCAAA 1878

RESULT 3
AL356315 158239 bp DNA HTG 22-MAR-2001
LOCUS Homo sapiens chromosome 1 clone RP11-332F5, *** SEQUENCING IN
DEFINITION
ACCESSION AL356315.8 GI:13445377
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 158239)
AUTHORS Plumb, B.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CU10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonequests@sanger.ac.uk
On Mar 24, 2001 this sequence version replaced gi:13443378.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba332F5

```


[illegible]


```

HSY17975      622 bp      mRNA      PRI      16-MAR-2000
LOCUS        HSY17975
DEFINITION   Homo sapiens mRNA for putative vacuolar ATPase membrane
sector associated protein M8-9.
ACCESSION    Y17975
VERSION      Y17975.1 GI:3451260
KEYWORDS     vacuolar proton ATPase membrane sector associated protein M8-9.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 622)
AUTHORS      Ludwig,J., Kerschner,S., Brandt,U., Pfeiffer,K., Getlawi,F.,
              Apps,D.K. and Schagger,H.
              Identification and characterization of a novel 9.2-kDa membrane
              sector-associated protein of vacuolar proton ATPase from chromaffin
              granules
JOURNAL      J. Biol. Chem. 273 (18), 10939-10947 (1998)
MEDLINE      98225166
REFERENCE    2 (bases 1 to 622)
AUTHORS      Ludwig,J.H.
              Direct Submission
              Submitted (20-AUG-1998) J.H. Ludwig, Universitaetsklinikum
              Frankfurt, Gustav Embden-Zentrum der Biologischen Chemie (ZBC),
              Institut fuer Biochemie I, Theodor Stern-Kai 7 Haus 25 B, 60590
              Frankfurt, FRG
              Related entry: R78826.
              Location/Qualifiers
                source          1..622
                               /organism="Homo sapiens"
                               /db_xref="taxon:9606"
                               192..494
                               /codon_start=1
                               /product="putative vacuolar proton ATPase membrane sector
                               associated protein M8-9"
                               /protein_id="CAI76984.1"
                               /db_xref="GI:3451261"
                               /db_xref="SPTREMBL:O75787"
                               /translation="MYSLVGNAVVELVTVKSFPTSLIRKTRTILEAKOAKNPASPYN
                               LAKRYRESVFENVNVLIMIALALAVITTSNTINMNDPRODSIIIRKMTNOKIRMD"
BASE COUNT   186 a      103 c      134 g      199 t
ORIGIN
Query Match      30.6%; Score 622; DB 97; Length 622;
Best Local Similarity 100.0%; Pred. No. 2.9e-122;
Matches 622; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

|||||
Db      361 TCACATGTCCTTGGATATGATCGCTTGGCTGTGTGATATATACCTTTACA 420
Qy      1076 atatttgaacatlgatccctlgatagatagatcatattatagatagacaacagaaga 1135
Db      421 ATATTGGACATGATCCGTGATATGATGATCATATTATATAGATGACCAACAGAGA 480
Qy      1136 ttcgaatgattgaatgattactctgctgccaagaattagaagaagggttgaattgctgt 1195
Db      481 TTGCATGATGATTAAGTTCCTGTCGTCAGATTAAGAAAAGGGGTTGGAAATGGCTGT 540
Qy      1196 ttgttaaatatatactctttagtgccttaagaatagatagatatacttaccattaa 1255
Db      541 TTGTAAATATATATCTTTACGTGCTTTAAAGTAGATATATCTTACATTATATA 600
Qy      1256 aaaaatcaaatltgttctt 1277
Db      601 AAAAATCAAAATTTGTCTTT 622

```

```

RESULT      8
LOCUS       G27225      513 bp      DNA
DEFINITION  human STS SHGC-31934, sequence tagged site.
ACCESSION   G27225
VERSION     G27225.1 GI:1375475
KEYWORDS    STS; STS sequence; primer; sequence tagged site.
SOURCE      human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 513)
REFERENCE    1 (bases 1 to 513)
AUTHORS      Myers,R.M.
              Unpublished (1995)
COMMENT      Contact: Richard M. Myers
              Stanford Human Genome Center (SHGC)
              Stanford University School of Medicine
              Department of Genetics, M-344, Stanford, CA 94305, USA
              Tel: 4157259687
              Fax: 4157259689
              Email: myers@shgc.stanford.edu

```

```

Primer A: CTTGATAGACATTTGCTCCACTCC
Primer B: TTTTATTCAGTAACTTTCCCCC
STS size: 128
PCR Profile:
Initial incubation: 94 degrees C for 90 seconds

```

```

Denaturation: 94 degrees C for 15 seconds
Annealing: 62 degrees C for 23 seconds
Polymerization: 72 degrees C for 30 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9600
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Tag Polymerase: 0.05 units/ul
Total Vol: 10 ul

```

```

Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 20 mM
pH: 8.3

```

Prepared with primer pairs provided by Sandoz, derived from R51314
-- Washington University/Merck EST sequence.

```

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

/map="x"
 STS 38.165
 primer_bind 38.60
 primer_bind complement(142..165)
 BASE COUNT 150 a 97 c 84 g 177 t 5 others
 ORIGIN

Query Match 18.6%; Score 378.6; DB 54; Length 513;
 Best Local Similarity 90.2%; Pred. No. 1.7e-70;
 Matches 458; Conservative 0; Mismatches 44; Indels 6; Gaps 5;

OY 1531 aggtgaatcatatttgggtctatgtaactgaacaaattgaatgcttaattg 1590
 DB 513 AGGNGGAAACCAATTTGGGCGCATGTAACCTAGGAACATGTAAGCCCTAAATGGA 454
 OY 1591 atgataaactcgaacaaagaagaagttttaa-cttaagtagcctaataatg 1649
 DB 453 AGTAATAACNCGGAAACAGAGAAAGTTTAACTAAGAGTACCCCAAAATATG 394
 OY 1650 gatgag-cttataatacgccttagtttggaactgatactgaagaaagagacg-ctg 1707
 DB 393 GATGTCCCTTAAATTAATGCGCTTGTGGAACGTATCCGAGTAACAGAGACAGCCTG 334
 OY 1708 ttcttaacccctctcgaagttgtagcc--tacaatggactaataatgatacctaa 1765
 DB 333 TTTTAAACCCCMTCGCAAGTTTGTGACCCCTACATGSGCCCTATATGATCTAAA 274
 OY 1766 atactacatgat-ctagaagaagaactagcctctggaagataatagatgcttcaat 1824
 DB 273 ATACTACATGATCTCTAAGAAAGAACTGACCTGTGAGATATATGATGCTTTCAATAT 214
 OY 1825 accacaaanaaccctgaaggaacatttgaagcactgataataaacatttattcagt 1884
 DB 213 ACACACAAAATCCCTGAGGAGCATTTTGAGCATGATAATAAACATTTTATTTCAGT 154
 OY 1885 aactttccctctgtgaagttacatggtttgtggaacacttcatcctagatatt 1944
 DB 153 AACTTTCCCTCTGTGAATGTAATGTTGTGAGCAACTCATTTCTATGAAATAT 94
 OY 1945 aagttgaagttggtgaattctactcttatttattgtgaagtgacaaatgctcacaagtg 2004
 DB 93 AAGTGAAGTGGTGAATTCCTACTTTTATGTGAGTGGACCATGCTCAAGTG 34
 OY 2005 acgaataaagttaagtatgattccaaaa 2032
 DB 33 ACAATAAAGTTAATGATGATTCACAAA 6

RESULT 9
 A74463 204 bp DNA PAT 15-OCT-1999
 LOCUS Sequence 149 from Patent WO9401548.
 DEFINITION A74463
 ACCESSION A74463.1 GI:6064477
 VERSION A74463.1
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 204)
 AUTHORS Sibson,D.R. and Gross,J.
 TITLE HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE,
 JOURNAL PLACENTA OR BONE MARROW
 MEDICAL RES COUNCIL (GB); SIBSON DAVID ROSS (GB)
 FEATURES
 Location/Qualifiers
 1..204
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 58 a 18 c 41 g 81 t 6 others
 ORIGIN

Query Match 8.8%; Score 179; DB 9; Length 204;
 Best Local Similarity 99.0%; Pred. No. 4.5e-28;
 Matches 190; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 1143 ggaattgaatgtaccctgtgcagaaattagaagaagggttgaattgctgtttgta 1202
 DB 14 GNAATGAATGTACCTGTGCCAGAAATAGAAAAGGGGTTGGAATTTGCTGTTTGTGA 73
 OY 1203 aaatatacttttagtgtgtttaaagtagtagtacttacttataaaaaaat 1262
 DB 74 AAATATATCTTTTGTGCTTTAAAGTAGATGATACTTTTAT-AAAAAAAT 132
 OY 1263 caaattgtcttctattttgtgtgtgcctgtatgttttctagaagtgatattagat 1322
 DB 133 CAAATTTGTTCTTTATTTTGTGTGCTGTGATGTTTCTACAGTAATATATGAT 192
 OY 1323 tgacgtgaatcc 1334
 DB 193 TGACGTGAATCC 204

RESULT 10
 A77442 204 bp DNA PAT 19-OCT-1999
 LOCUS Sequence 149 from Patent EP0587279.
 DEFINITION A77442
 ACCESSION A77442
 VERSION A77442.1 GI:6089107
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 204)
 AUTHORS Sibson,D.R. and Hadfield,K.M.
 TITLE HUMAN NUCLEIC ACID FRAGMENTS ISOLATED FROM BRAIN, ADRENAL TISSUE,
 JOURNAL PLACENTA OR BONE MARROW AND THEIR USE
 MEDICAL RES COUNCIL (GB)
 FEATURES
 Location/Qualifiers
 1..204
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 58 a 18 c 41 g 81 t 6 others
 ORIGIN

Query Match 8.8%; Score 179; DB 9; Length 204;
 Best Local Similarity 99.0%; Pred. No. 4.5e-28;
 Matches 190; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 1143 ggaattgaatgtaccctgtgcagaaattagaagaagggttgaattgctgtttgta 1202
 DB 14 GNAATGAATGTACCTGTGCCAGAAATAGAAAAGGGGTTGGAATTTGCTGTTTGTGA 73
 OY 1203 aaatatacttttagtgtgtttaaagtagtagtacttacttataaaaaaat 1262
 DB 74 AAATATATCTTTTGTGCTTTAAAGTAGATGATACTTTTAT-AAAAAAAT 132
 OY 1263 caaattgtcttctattttgtgtgtgcctgtatgttttctagaagtgatattagat 1322
 DB 133 CAAATTTGTTCTTTATTTTGTGTGCTGTGATGTTTCTACAGTAATATATGAT 192
 OY 1323 tgacgtgaatcc 1334
 DB 193 TGACGTGAATCC 204

RESULT 11
 AC026156/c 126543 bp DNA HMG 28-MAR-2000
 LOCUS AC026156
 DEFINITION Homo sapiens chromosome 3p clone RP11-169016, WORKING DRAFT
 SEQUENCE, 10 unordered pieces.
 ACCESSION AC026156
 VERSION AC026156.2 GI:7331308

KEYWORDS HTGS_PHASE1; HTGS_DRAFT.

SOURCE human

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 126543)

AUTHORS Lu, G., Zhao, Y., Lin, W., Dong, H., Wan, M., Xu, S., Gu, W., Tu, Y., Jia, J., Wu, C., Zhang, C., Zhong, M., Zhou, Y., Ren, S., Fu, G., Chen, Z. and Huang, M.

TITLE Direct Submission

JOURNAL Submitted (21-MAR-2000) Genomic Dept., Chinese National Human Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai, Shanghai 201203, P. R. China

COMMENT On Mar 26, 2000 this sequence version replaced gi:7271973.

Center: Chinese National Human Genome Center at Shanghai

Center Code: CHGC

Web site: <http://www.chgc.sh.cn>

Email: mhuang@chgc.sh.cn or fugang@chgc.sh.cn

-----End Genome Center Information-----

* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 3411: contig of 3411 bp in length
* gap of unknown length
* 3412 7987: contig of 4576 bp in length
* gap of unknown length
* 7988 16490: contig of 8503 bp in length
* gap of unknown length
* 16491 30218: contig of 13728 bp in length
* gap of unknown length
* 30219 44895: contig of 14677 bp in length
* gap of unknown length
* 44896 55628: contig of 10733 bp in length
* gap of unknown length
* 55629 71833: contig of 16205 bp in length
* gap of unknown length
* 71834 84537: contig of 12704 bp in length
* gap of unknown length
* 84538 103706: contig of 19169 bp in length
* gap of unknown length
* 103707 126543: contig of 22837 bp in length.

FEATURES Location/Qualifiers

1..126543 Homo sapiens

/db_xref="taxon:9606"

/chromosome="3p"

/clone="RP11-169016"

BASE COUNT 34774 a 28466 c 28425 g 34834 t 44 others

ORIGIN

Query Match 7.58: Score 152; DB 69; Length 126543;
Best Local Similarity 94.0%; Pred. No. 4; Be-22;
Matches 156; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 669 tgatattcaagcttgctgcgtcgcataagcatcacaagatcatctcctgattt 728

DB 89824 TGTGTTGTTCTCTAAGCTGCTGCTCTCATAGCATCTACCCAGGATCATCTCCGATTT 89765

QY 729 atattcactgagctgagcaggcttgatcgaatcgggaagcgttatagggaagactctga 788

DB 89764 ATATTCACTGAGCTGCGAGCTTGGATGAATTTGGCAAGCGATTATGGGAAGACTCTGA 89705

QY 789 acaattcagagatgctcttaagatcctttgaagctctgcaaaagt 836

DB 89704 ACAATTGAGAGATGCTCTCAAGATCTGTGTACCGCTCTGCAAAAGCT 89657

RESULT 12

AF259074/c

LOCUS AF259074 72831 bp DNA ROD 21-JUN-2000

DEFINITION Mus musculus T-cell receptor alpha locus BAC clone MBAC519 from 14D1-D2, complete sequence.

ACCESSION AF259074

VERSION AF259074.1 GI:8575574

KEYWORDS

SOURCE

ORGANISM

house mouse

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished
2 (bases 1 to 72831)
Lee, I.Y., Wang, K., Lasky, S.R., Dahl, T., Hall, J. and Hood, L.E.
Direct Submission
Submitted (21-APR-2000) Department of Molecular Biotechnology, Box 357730 University of Washington, Seattle, WA 98195, USA
Interpersed Repeats were identified with RepeatMasker (available from <http://ftp.genome.washington.edu/RM/RepeatMasker.html>) Simple sequence repeats were identified with spunk (available from <http://serac.mbt.washington.edu/chrisa/software/spunk.html>).

FEATURES Location/Qualifiers

1..72831

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="14"

/map="14D1-D2"

/clone="MBAC519; RG:298L19"

/cell_line="CJ7"

/clone_lib="Research Genetics Mouse BAC Library"

/note="T-cell receptor alpha locus; flanking sequences are UM:MBAC1058 (Genbank Accession Number AF259072) 5' and UM:MBAC01 (Genbank Accession Number AF259071) 3'.

complement(1..3813)

/rpt_family="L1_MM"

complement(3794..9677)

/rpt_family="L1_MM"

complement(9585..10193)

/rpt_family="L1F_Send"

complement(9922..10405)

/rpt_family="L1F_Send"

complement(10134..10617)

/rpt_family="L1F_Send"

complement(10346..10666)

/rpt_family="L1F_Send"

complement(10688..11092)

/rpt_family="L1M2_or12"

complement(11099..11689)

/rpt_family="L1_RN"

11945..11972

/rpt_family="(CAT)n"

12294..12490

/rpt_family="RLTR13D"

13228..13278

/rpt_family="(TGG)n"

13291..13464

/rpt_family="(TGG)n"

14374..15028

/rpt_family="RLTR13C"

14515..15028

/rpt_family="RLTR13D"

15157..15862

/rpt_family="L1_MM"

15873..16356

/rpt_family="L1F_Send"

16085..16568

/rpt_family="L1F_Send"

repeat_region	16509 . 16992 /rpt_family="L1F_Send"
repeat_region	16933 . 17541 /rpt_family="L1F_Send"
repeat_region	17449 . 23342 /rpt_family="L1_MM"
repeat_region	23348 . 23928 /rpt_family="L1_MM"
repeat_region	23365 . 24582 /rpt_family="L1_RN"
repeat_region	24564 . 24701 /rpt_family="L1_MM"
repeat_region	24599 . 26220 /rpt_family="L1_MM"
repeat_region	complement(26217 . 27218) /rpt_family="L1_MM"
repeat_region	complement(27118 . 27357) /rpt_family="L1F_Send"
repeat_region	27881 . 28092 /rpt_family="B3"
repeat_region	complement(28543 . 28587 /rpt_family="(CA)n"
repeat_region	complement(28929 . 28969) /rpt_family="(CA)n"
repeat_region	28938 . 28969 /rpt_unit="(Gr)x15
repeat_region	29384 . 29531 /rpt_family="MTD"
repeat_region	29536 . 29866 /rpt_family="L1_RN"
repeat_region	29875 . 30280 /rpt_family="RLTR13C"
repeat_region	30282 . 31643 /rpt_family="L1_MM"
repeat_region	31674 . 33181 /rpt_family="RLTR13C"
repeat_region	32186 . 34760 /rpt_family="L1_MM"
repeat_region	34747 . 36147 /rpt_family="L1_RN"
repeat_region	34929 . 34953 /rpt_unit="(AAMC)x6
repeat_region	34930 . 34953 /rpt_family="(CAA)n"
repeat_region	36156 . 36306 /rpt_family="Lx3"
repeat_region	36337 . 36357 /rpt_family="RMER13A"
repeat_region	36561 . 37025 /rpt_family="RMER13A"
repeat_region	37029 . 37841 /rpt_family="Lx4"
repeat_region	37821 . 38430 /rpt_family="Lx4"
repeat_region	complement(38431 . 38476) /rpt_family="AT-rich"
repeat_region	39113 . 39169 /rpt_unit="(CA)x28
repeat_region	39114 . 39169 /rpt_family="(CA)n"
repeat_region	complement(39309 . 40536) /rpt_family="I1_MM"
repeat_region	complement(40537 . 40629) /rpt_family="Lx5"
repeat_region	40636 . 40687 /rpt_family="RMER6A"
repeat_region	40688 . 41215 /rpt_family="RMER6A"
repeat_region	complement(41335 . 41441) /rpt_family="RMER1B"
repeat_region	41452 . 41573 /rpt_family="ORR1A3"
repeat_region	42854 . 42994

```
repeatL_region /rpt_family="B1-F" 42986, .43037
/rpt_family="CAAAn"
repeatL_region /rpt_family="CAAAn" complement(44550, .44693)
/rpt_family="RMR1B"
repeatL_region /rpt_family="RMR1B" complement(44699, .44893)
/rpt_family="RMR1B"
repeatL_region 44997, .45152
/rpt_family="MERVL"
repeatL_region 45066, .45164
/rpt_family="HERVL"
repeatL_region /rpt_family="HERVL" 45202, .46018
/rpt_family="MERVL"
repeatL_region /rpt_family="MERVL" complement(46029, .46760)
/rpt_family="L1_MM"
repeatL_region 46762, .46985
/rpt_family="MERVL"
repeatL_region complement(46986, .47171)
/rpt_family="ORR1A1"
repeatL_region /rpt_family="ORR1A1" complement(47015, .47353)
/rpt_family="ORR1C"
repeatL_region 47201, .47258
/rpt_family="ORR1C"
repeatL_region /rpt_unit=(AG)x28 47202, .47258
/rpt_family="GA)n"
repeatL_region complement(47350, .47483)
/rpt_family="RMR1B"
repeatL_region complement(47643, .48403)
/rpt_family="RMR1B"
repeatL_region complement(48459, .48558)
/rpt_family="ORR1C"
repeatL_region 48573, .48751
/rpt_family="MERVL"
repeatL_region 48611, .48933
/rpt_family="HERVL"
repeatL_region 48940, .49259
/rpt_family="L1_MM"
repeatL_region 49286, .49549
/rpt_family="MERVL"
repeatL_region complement(49570, .49703)
/rpt_family="(7A)n"
repeatL_region 49751, .49830
/rpt_family="B1-F"
repeatL_region complement(49884, .50083)
/rpt_family="L1ME2"
repeatL_region complement(49885, .50110)
/rpt_family="L1ME1"
repeatL_region 51858, .51878
/rpt_family="AT_Rich"
repeatL_region 52211, .52307
/rpt_family="PBID9"
repeatL_region 52224, .52294
/rpt_family="B1_MM"
repeatL_region 52319, .52422
/rpt_family="L1ME2_Ort2"
repeatL_region 52424, .52487
/rpt_family="B1-F"
```

	Query Match	6.5%	Score 132.2	DB 94	Length 72831
	Best Local Similarity	91.5%	Pred. No. 7.4e-18		
	Matches 140	Conservative	0	Mismatches 13	Indels 0
				Gaps 0	
QY	684	gcctgcgcgtcaaaagactctgaagcaagatcatctccctggttaataatcactgaagct	743		
Db	41738	gtttgcttcgcataaagcattcaccgacagaccattctcccgactgtttatctatggaact	41679		
QY	744	ggcagcgtttgagatgaatttggagacgctatcaggggaagactctgaacaattcagaatgc	803		
Db	41678	ggcagcgtttgagatgaacattggggaagcgctttatcaggaagacctctgaacagttcaggatgc	41619		
QY	804	ttcctaagatcccttgttgaagcgtctgcaaaagt	836		
Db	41618	ttcttaacatccctgttgtagctctcccaaaagct	41586		

```

RESULT 13
A74460 151 bp DNA PAT 15-OCT-1999
DEFINITION Sequence 146 from Patent WO9401548.
ACCESSION A74460
VERSION A74460.1 GI:6064474
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 151)
AUTHORS Sibson,D.R. and Gross,J.
TITLE HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE,
JOURNAL PLACENTA OR BONE MARROW
MEDICAL RES COUNCIL (GB); SIBSON DAVID ROSS (GB)
FEATURES
Location/Qualifiers
source 1.151
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 49 a 14 c 25 g 58 t 5 others
ORIGIN

Query Match 6.1%; Score 123.4; DB 9; Length 151;
Best Local Similarity 98.5%; Pred. No. 2.9e-16;
Matches 135; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1145 atggaatgttaccctgtgcagaaatagaagaagggtgtgaagaattgctgtttgtttaa 1204
Db 16 ATTGAATGTTACCTGTGCCAGAAATTAGAAAAGGGGTTGGAAATTGGCTGTTTGTAAAA 75

Oy 1205 atatacttttagtgtgtcttaagtagatagatattacattacattataaaaaaatca 1264
Db 76 ATATATCTTTAGTGTCTTTAAAGTAGATAGTATACCTTTACATTTAT-AAAAAATCA 134

Oy 1265 aattgttcttattt 1281
Db 135 AATTTTGTTCTTTAATT 151

RESULT 14
A77439 151 bp DNA PAT 19-OCT-1999
DEFINITION Sequence 146 from Patent EP0587279.
ACCESSION A77439
VERSION A77439.1 GI:6089104
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 151)
AUTHORS Sibson,D.R. and Hadfield,K.M.
TITLE HUMAN NUCLEIC ACID FRAGMENTS ISOLATED FROM BRAIN, ADRENAL TISSUE,
JOURNAL PLACENTA OR BONE MARROW AND THEIR USE
MEDICAL RES COUNCIL (GB)
FEATURES
Location/Qualifiers
source 1.151
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 49 a 14 c 25 g 58 t 5 others
ORIGIN

Query Match 6.1%; Score 123.4; DB 9; Length 151;
Best Local Similarity 98.5%; Pred. No. 2.9e-16;
Matches 135; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1145 atggaatgttaccctgtgcagaaatagaagaagggtgtgaagaattgctgtttgtttaa 1204

```

```

Db 16 ATTGAATGTTACCTGTGCCAGAAATTAGAAAAGGGGTTGGAAATTGGCTGTTTGTAAAA 75

Oy 1205 atatacttttagtgtgtcttaagtagatagatattacattacattataaaaaaatca 1264
Db 76 ATATATCTTTAGTGTCTTTAAAGTAGATAGTATACCTTTACATTTAT-AAAAAATCA 134

Oy 1265 aattgttcttattt 1281
Db 135 AATTTTGTTCTTTAATT 151

RESULT 15
A74462 148 bp DNA PAT 15-OCT-1999
DEFINITION Sequence 148 from Patent WO9401548.
ACCESSION A74462
VERSION A74462.1 GI:6064476
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 148)
AUTHORS Sibson,D.R. and Gross,J.
TITLE HUMAN NUCLEIC ACID FRAGMENTS, ISOLATED FROM BRAIN ADRENAL TISSUE,
JOURNAL PLACENTA OR BONE MARROW
MEDICAL RES COUNCIL (GB); SIBSON DAVID ROSS (GB)
FEATURES
Location/Qualifiers
source 1.148
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 47 a 14 c 25 g 57 t 5 others
ORIGIN

Query Match 6.0%; Score 122.4; DB 9; Length 148;
Best Local Similarity 98.5%; Pred. No. 4.7e-16;
Matches 134; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1145 atggaatgttaccctgtgcagaaatagaagaagggtgtgaagaattgctgtttgtttaa 1204
Db 14 ATTGAATGTTACCTGTGCCAGAAATTAGAAAAGGGGTTGGAAATTGGCTGTTTGTAAAA 73

Oy 1205 atatacttttagtgtgtcttaagtagatagatattacattacattataaaaaaatca 1264
Db 74 ATATATCTTTAGTGTCTTTAAAGTAGATAGTATACCTTTACATTTAT-AAAAAATCA 132

Oy 1265 aattgttcttattt 1280
Db 133 AATTTTGTTCTTTAATT 148

```

Search completed: August 18, 2001, 22:16:45
 Job time: 14556 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: August 18, 2001, 21:50:38 ; Search time 250.24 Seconds

(without alignments)
5101.199 Million cell updates/sec

Title: US-09-284-320-56

Perfect score: 2033

Sequence: 1 gagtcgagcgccgcaccc.....gttaatgatgatcccaaac 2033

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq.0601:*
1: /cgnl_9/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /cgnl_9/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /cgnl_9/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /cgnl_9/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /cgnl_9/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /cgnl_9/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /cgnl_9/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /cgnl_9/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /cgnl_9/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /cgnl_9/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /cgnl_9/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /cgnl_9/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /cgnl_9/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /cgnl_9/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /cgnl_9/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /cgnl_9/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /cgnl_9/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /cgnl_9/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /cgnl_9/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /cgnl_9/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /cgnl_9/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /cgnl_9/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2033	100.0	2033	19	AAV93560 Human epidermal ca
2	2021.6	99.4	2092	22	AAF93774 Human CDNA exocodin
3	2020	99.4	2108	22	AAF97893 Human secreted pro
4	1987.6	97.8	2099	22	AAF97929 Human secreted pro
5	1914.8	94.2	2010	20	AAZ40848 Secreted protein E
6	1050	51.6	1050	19	AAV49561 Human epidermal ca
7	639	31.4	793	20	AAZ17313 Human gene express
8	639	31.4	793	20	AAV98788 Human validated ca
9	547.2	26.9	577	20	AAV87120 EST clone BOB7. H
10	547	26.9	813	20	AAZ15750 Human gene express
11	547	26.9	813	20	AAZ15759 Human gene express

C	12	512.6	25.2	588	22	AAF94119	Primer specific fo
C	13	486.8	23.9	560	22	AAF93970	Primer specific fo
C	14	386.2	19.0	858	20	AAZ15793	Human gene express
C	15	371.2	18.3	416	20	AAZ17715	Human gene express
C	16	322	15.8	343	20	AAV51918	Human secreted pro
C	17	277.8	13.7	828	20	AAZ15908	Human gene express
C	18	255	12.5	260	16	AAV20085	Human gene signatu
C	19	244	12.0	297	20	AAZ13071	Human gene express
C	20	192.6	9.5	936	22	AAF58252	Oligonucleotide D1
C	21	192.6	9.5	936	22	AAF58254	Oligonucleotide D1
C	22	192.6	9.5	936	22	AAF58257	Oligonucleotide D1
C	23	192.6	9.5	936	22	AAF58259	Oligonucleotide D2
C	24	192.6	9.5	936	22	AAF58262	Oligonucleotide D2
C	25	192.6	9.5	936	22	AAF58255	Oligonucleotide D1
C	26	186.6	9.2	936	22	AAF58252	Oligonucleotide D1
C	27	186.6	9.2	936	22	AAF58254	Oligonucleotide D1
C	28	186.6	9.2	936	22	AAF58257	Oligonucleotide D1
C	29	186.6	9.2	936	22	AAF58259	Oligonucleotide D2
C	30	186.6	9.2	936	22	AAF58262	Oligonucleotide D2
C	31	186.6	9.2	936	22	AAF58255	Oligonucleotide D1
C	32	179	8.8	204	15	AAV76549	Human genome fragm
C	33	166.2	8.2	254	20	AAZ12587	Human genome fragm
C	34	153.8	7.6	469	21	AAV43203	Xenopus secreted e
C	35	123.4	6.1	151	15	AAV76546	Human genome fragm
C	36	120.8	5.9	148	15	AAV76548	Human genome fragm
C	37	67.4	3.3	244	22	AAF58238	Oligonucleotide D1
C	38	62	3.0	244	22	AAF58238	Oligonucleotide D1
C	39	51.4	2.5	1864	8	AAV71405	Sequence of ANS-1
C	40	49.2	2.4	2781	21	AAV87991	Kapok CPB-1A 5' u
C	41	47.4	2.3	6152	18	AAV78867	P. falciparum live
C	42	46.2	2.3	418	21	AAV94060	Cat flea hindgut a
C	43	46	2.3	422	21	AAV94573	Cat flea hindgut a
C	44	46	2.3	1864	15	AAV78892	Aspergillus nidula
C	45	45.6	2.2	19124	18	AAV72882	Plasmodium var-7 g

ALIGNMENTS

RESULT	1
ID	AAV49560 standard; CDNA to mRNA: 2033 BP.
XX	AAV49560:
AC	AAV49560: (first entry)
XX	
DT	21-OCT-1998
XX	
DE	Human epidermal carcinoma cell line KB clone HP01293 CDNA #1.
XX	
KW	Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW	differentiation; immune system; stimulator; suppressor; regulator;
KW	haematopoiesis; activin; inhibitor; chemotactic; chemokine; receptor;
KW	haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX	
OS	Homo sapiens.
XX	
EH	Key
FT	97.1149
FT	Location/Qualifiers
FT	/tag= a
FT	/product= "transmembrane domain containing protein"
XX	
PN	W09821328-A2.
XX	
PD	22-MAY-1998.
XX	
PF	07-NOV-1997: 97WO-JP04056.
XX	
PR	13-NOV-1996: 96JP-0301429.
XX	
PA	(PROT-) PROTEGNE INC.
PA	(SAGA) SAGAMI CHEM RES CENTRE.
XX	
PI	Kato S, Kobayashi M, Sekine S, Yamaguchi T;

Db	1801	gagctatagatgcctttccattacacacaaaatcccgaggagacattttgagcagc	1860
Qy	1861	aataaacaattttatctttagtaacttttcccccgtglaagttactatggttttgg	1920
Db	1861	aataaacaattttatctttagtaacttttcccccgtglaagttactatggttttgg	1920
Qy	1921	tacaacttcattcatagaatataatagtggaagtgggtgaattctacttttaagtgtga	1980
Db	1921	tacaacttcattcatagaatataatagtggaagtgggtgaattctacttttaagtgtga	1980
Qy	1981	gtggagccaatgctctatcaaggtggaacaataatgaatgaatgattccaaac	2033
Db	1981	gtggagccaatgctctatcaaggtggaacaataatgaatgaatgattccaaac	2033

RESULT 2

AAFP3774	ID	AAFP3774 standard; cDNA; 2092 BP.
XX	AC	
XX	AAFP3774;	
XX	DT	23-MAY-2001 (first entry)
XX	DE	Human cDNA encoding a membrane or secretory protein clone PSEC0072.
XX	OS	Human; secretory protein; membrane protein; vaccine; gene therapy; rheumatoid arthritis; diabetes; ss.
XX	XX	Homo sapiens.
XX	XX	BP1067182-A2.
XX	XX	10-JAN-2001.
XX	XX	07-JUL-2000; 2000EP-0114090.
XX	XX	08-JUL-1999; 99JP-0194179.
XX	XX	11-JAN-2000; 2000JP-0118775.
XX	XX	02-MAY-2000; 2000JP-0183766.
XX	XX	(HELI-) HELIX RES INST.
XX	XX	Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX	XX	WPI; 2001-093989/11.
XX	XX	P-PSDB; AAB88347.
XX	XX	Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development -
XX	XX	Claim 1; SEQ ID 61; 609pp + CD ROM; English.
XX	XX	
XX	XX	This invention relates to nucleic acid sequences AAFP3774 - AAFP3916
XX	XX	which encode human secretory or membrane proteins represented by
XX	XX	AAB88317 - AAB88419. Included in the invention are primers
XX	XX	AAFP3917 - AAFP4295 and AAFP6223 - AAFP6235 which are used to isolate the
XX	XX	cDNA sequences of the invention. The invention also includes methods for
XX	XX	the production of antibodies directed against the proteins, and cDNA
XX	XX	sequences, which can be used in vaccines. The polynucleotide sequences
XX	XX	can be used in gene therapy. The polynucleotide sequences and the
XX	XX	proteins they encode may be used in the prevention, treatment and
XX	XX	diagnosis of diseases associated with inappropriate secretory
XX	XX	protein/membrane protein expression. The nucleic acids and complementary
XX	XX	sequences may also be used as DNA probes in diagnostic assays
XX	XX	(e.g. polymerase chain reactions (PCR)) to detect and quantify the
XX	XX	presence of similar nucleic acid sequences in samples. They may also be
XX	XX	used to study the expression and function of secretory proteins/membrane
XX	XX	polypeptides and their role in metabolism. The polypeptides may be used
XX	XX	as antigens in the production of antibodies against them and in assays to
XX	XX	identify modulators (agonists and antagonists) of expression and
XX	XX	activity. The antibodies and antagonists may also be used as therapeutic
XX	XX	agents to down regulate expression and activity. The antibodies may also
XX	XX	be used as diagnostic agents for detecting the presence of the
XX	XX	

Dh 965 ttgacacctcccttaaggaagacagactactcttgaggcaaacagcgaagac 1024
OY 961 ccaggcaagcccttaaccccttgatataagataatttgaatattccgtgtttcaac 1020
Dh 1025 ccaggcaagcccttaaccccttgatataagataatttgaatattccgtgtttcaac 1084
OY 1021 atgtaacttgataatgatacgcttgcccttgctgtgtatatacactcttaacaatatt 1080
Dh 1085 atgtactcttgataatgatacgcttgcccttgctgtgtatatacactcttaacaatatt 1144
OY 1081 tggaaatagatcccttgatatagtatagatcttattatggatgagcaaaccaaaatattcga 1140
Dh 1145 tggaaatagatcccttgatatagtatagatcttattatggatgagcaaaccaaaatattcga 1204
OY 1141 atgtaacttgatcttactctgtgcgaagatagaagagggttctgaattgctgtttgtg 1200
Dh 1205 atgtaacttgatcttactctgtgcgaagatagaagagggttctgaattgctgtttgtg 1264
OY 1201 taaatatactctttagtgccttaagtagatagatatacttaccattataaaaaaa 1260
Dh 1265 taaatatactctttagtgccttaagtagatagatatacttaccattataaaaaaa 1324
OY 1261 atcaaatctgtcttcttatttctgtgtgtgcctgtgacgtttctctagaatgaattatagt 1320
Dh 1325 atcaaatctgtcttcttatttctgtgtgtgcctgtgacgtttctctagaatgaattatagt 1384
OY 1321 atgaagctgaatcccaactgtgfatagattccataatatacttgatattatagatag 1380
Dh 1385 atgaagctgaatcccaactgtgfatagattccataatatacttgatattatagatag 1444
OY 1381 catttaatacaatctgatttcaatctgtttaaagaatttgaagaatctgacgtgaagaagaat 1440
Dh 1445 catttaatacaatctgatttcaatctgtttaaagaatttgaagaatctgacgtgaagaagaat 1504
OY 1441 gtaaaacatttagaagatgcctgcgttatggaaaaaagtgcacgaattatataaacaac 1500
Dh 1505 gtaaaacatttagaagatgcctgcgttatggaaaaaagtgcacgaattatataaacaac 1564
OY 1501 taagaatgcttaactcttcttaacacagatagtgtaaaatcatatttggctatgtata 1560
Dh 1565 ttacgaatgcttaactcttcttaacacagatagtgtaaaatcatatttggctatgtata 1624
OY 1561 ctatgaacaatttctaatttctaatttgaatgtaaaatcttaacaagaagaagaag 1620
Dh 1625 ctatgaacaatttctaatttctaatttgaatgtaaaatcttaacaagaagaagaag 1684
OY 1621 ttttaacttaagtagcccttaaaatagatgtgcttatataatcogcttagtttggaa 1680
Dh 1685 ttttaacttaagtagcccttaaaatagatgtgcttatataatcogcttagtttggaa 1744
OY 1681 ctgtatcgtgaatacagagagacagctgttttttaacccctcttcaagtttgttagcct 1740
Dh 1745 ctgtatcgtgaatacagagagacagctgttttttaacccctcttcaagtttgttagcct 1804
OY 1741 acatgggctaataatgataactaaataactacatctgatacgaagaagaactagcctgtg 1800
Dh 1805 acatgggctaataatgataactaaataactacatctgatacgaagaagaactagcctgtg 1864
OY 1801 gatatataagatgcttcttcaataacacacaaatcccgaggagacatttggaggaag 1860
Dh 1865 gatatataagatgcttcttcaataacacacaaatcccgaggagacatttggaggaag 1924
OY 1861 aataataaacatttattatcttaagtaacttcccccgtgtgaatgataatgattgtg 1920
Dh 1925 aataataaacatttattatcttaagtaacttcccccgtgtgaatgataatgattgtg 1984
OY 1921 taacaattcatctatagaataatgaatggagagtggttgatcttaatttattgttga 1980
Dh 1985 taacaattcatctatagaataatgaatggagagtggttgatcttaatttattgttga 2044
OY 1981 gtgagaccaatgtctataagaagtgcacaaataaagttaatgaatattcc 2028
Dh 2045 gtgagaccaatgtctataagaagtgcacaaataaagttaatgaatattcc 2092

RESULT 3
AAFP97893
ID AAFP97893 standard; cDNA; 2108 BP.
XX
AC AAFP97893;
DF 01-JUN-2001 (first entry)
DE Human secreted protein cDNA, SEQ ID NO: 20.
XX
KW Human; secreted protein; immunomodulatory; antisclerotic;
KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KW anticonvulsant; antialzheimer's; antiparkinsonian;
KW antimicrobial; vulnereary; vaccine; gene therapy; cancer;
KW protein coordinate data; infection; ss.
XX
OS Homo sapiens.
XX
PN WO200121658-A1.
PD 29-MAR-2001.
XX
PP 22-SEP-2000; 2000WO-US26013.
XX
PR 24-SEP-1999; 99US-0155709.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI NI J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
PI Young PE, Wei P, Florence KA;
XX
DR WPI; 2001-235311/24.
XX
PT Nucleic acids encoding 32 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Claim 1: Page 726-727; 890pp; English.
XX
CC The present sequence encodes one of 32 novel human secreted polypeptides.
CC The nucleic acid molecules and polypeptides they encode may be used in
CC the prevention, diagnosis and treatment of diseases such as
CC immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
CC and human immuno-deficiency virus (HIV) infections), hyperproliferative
CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC neovascularisation and diabetic retinopathy), neurological disorders
CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC infectious diseases and/or for promoting wound healing, regeneration
CC and/or chemotaxis. The nucleic acid molecules may be used to produce the
CC secreted polypeptides. They may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acid
CC sequences in samples. The polypeptides may be used as antigens in the
CC production of antibodies and in assays to identify modulators of
CC their expression and activity.
XX
SQ Sequence 2108 BP; 592 A; 389 C; 460 G; 664 T; 3 other:

Query Match 99.4%; Score 2020; DB 22; Length 2108;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 gtacactctcaagcgtgcgtgtgtgcgcgtgtccgcgcgcgttcgcgtgtgcgcgcg 72
Dh 46 gtacactctcaagcgtgcgtgtgtgtgcgcgtgtccgcgcgcgttcgcgtgtgcgcgcg 105
OY 73 cagtgctgcgcgcgcgcgcgcacatgctgtgttctgtcgtcctcgcgttcgtgtgcgcg 132

106 cagtgctgcgcccgcgcgcaccacgctgctgttgcgtcctcgtgcgttgcgtgcg 165
133 ggtgtcttggaagacgagtttaagatatataaataccacgagctgtgttctccgaat 192
166 ggtgtcttggaagacgagtttaagatatataaataccacgagctgtgttctccgaat 225
193 ggaatttgccataccagagagcgagccacgagctgtgtcatatgccaatggtctc 252
226 ggaatttgccataccagagagcgagccacgagctgtgtcatatgccaatggtctc 285
253 tctgtgaagaagacaccttcttgcgcagacgctgcagctgggttaacctgttcaatcgtc 312
286 tctgtgaagaagacaccttcttgcgcagacgctgcagctgggttaacctgttcaatcgtc 345
313 cgggtcaccgtaatggtgagtggaagggagtggaacaaactggtcttaaccccccgagct 372
346 cgggtcaccgtaatggtgagtggaagggagtggaacaaactggtcttaaccccccgagct 405
373 gtaatttcgtaaccttggaagatgcaagcttcccttgaagcttgacaagcttgacaatc 432
406 gtaatttcgtaaccttggaagatgcaagcttcccttgaagcttgacaagcttgacaatc 465
433 attcaactcttatcttctgcggaagactcctgttcttgcaagcttgccacgtaagaa 492
466 attcaactcttatcttctgcggaagactcctgttcttgcaagcttgccacgtaagaa 525
493 agagtgataatgtagaaggaagcaaacatgagtttgaagaaacttcaacacacacacacac 552
526 agagtgataatgtagaaggaagcaaacatgagtttgaagaaacttcaacacacacacacac 585
553 cagctccgtaacacgctgttctcaagaacactgttctcagttcaatccccctcaatctc 612
586 cagctccgtaacacgctgttctcaagaacactgttctcagttcaatccccctcaatctc 645
613 ctgagtagaagaacatgaaagtgcacgctccttcttctgaacatgcaaatgtaacatg 672
646 ctgagtagaagaacatgaaagtgcacgctccttcttctgaacatgcaaatgtaacatg 705
673 atttcaagcttctgtctcctcctaaagacatcagacgaagagatcattctcctatattat 732
706 atttcaagcttctgtcctcctcctaaagacatcagacgaagagatcattctcctatattat 765
733 tcaactgagacgtagcaggttctgagatgaaatttggaagcggtatggaagaaactcgaacaa 792
766 tcaactgagacgtagcaggttctgagatgaaatttggaagcggtatggaagaaactcgaacaa 825
793 ttcaagagatgcttcaagatccttctgagcgtctgcgaagaagtttgcagatacagctac 852
826 ttcaagagatgcttcaagatccttctgagcgtctgcgaagaagtttgcagatacagctac 885
853 agctcttaatggtggaatgcaatgtagagttagctacgttcaagttcaattgacacctcc 912
886 agctcttaatggtggaatgcaatgtagagttagctacgttcaagttcaattgacacctcc 945
913 ctcaattagaagaagaagacatccttgcgaagaagaagaagaagaagaagaagaagaaga 972
946 ctcaattagaagaagaagacatccttgcgaagaagaagaagaagaagaagaagaagaaga 1005
973 tataaccttgacataagatataatttgataatctcgtgtgttctcaacatggtactgtg 1032
1006 tataaccttgacataagatataatttgataatctcgtgtgttctcaacatggtactgtg 1065
1033 ataattgacgcttgcgtgctgagatatacactcttcaacatatttgagaactgagat 1092
1066 ataattgacgcttgcgtgctgagatatacactcttcaacatatttgagaactgagat 1125
1093 cctgagataatgataatcattatagatgacaacacgaagaatcgaatgagatgagatg 1152
1126 cctgagataatgataatcattatagatgacaacacgaagaatcgaatgagatgagatg 1185
1153 ttaccctgcgaagaattagaagaaggtggaatctgagtttggatataaataatct 1212
|||||

1186 ttaacctgcgacgaattagaagaaggtgtggaattgagctgttctgttaataataatct 1245
1213 tttaagtgctttaaagtagatagatatacttaacttaataaaaaaatacaatttgt 1272
1246 tttaagtgctttaaagtagatagatatacttaacttaataaaaaaatacaatttgt 1305
1273 tcttaattgtgtgtcctgtgagtggttctcgaagtgaaattatagatgagttgaat 1332
1306 tcttaattgtgtgtcctgtgagtggttctcgaagtgaaattatagatgagttgaat 1365
1333 cccaatggtgataatgataataatgcttgaaatattatgatatagccatttaataaa 1392
1366 cccaatggtgataatgataataatgcttgaaatattatgatatagccatttaataaa 1425
1393 ttgatttcaatttgtttaagaaatttggaatatgcaatgaaagaatgtaaaacttta 1452
1426 ttgatttcaatttgtttaagaaatttggaatatgcaatgaaagaatgtaaaacttta 1485
1453 gaatagctcgtgtatggaanaaagtgcacgtgaatttattagaacaaacttaagact 1512
1486 gaatagctcgtgtatggaanaaagtgcacgtgaatttattagaacaaacttaagact 1545
1513 aacttcttaacacgaatgagtgaaatcataatttggcgtattgtatatacgaacatt 1572
1546 aacttcttaacacgaatgagtgaaatcataatttggcgtattgtatatacgaacatt 1605
1573 tgttaatgtcttaatttgaatgataaataacttgaaacacgaagaagaagttttaaact 1632
1606 tgttaatgtcttaatttgaatgataaataacttgaaacacgaagaagaagttttaaact 1665
1633 agtagcccttaaaatagatgagtgcttataatacgttaatttggaaactatcagat 1692
1666 agtagcccttaaaatagatgagtgcttataatacgttaatttggaaactatcagat 1725
1693 aacagagagacgctgttctttaaaccctctctgcagagttgttgacacacacacacac 1752
1726 aacagagagacgctgttctttaaaccctctctgcagagttgttgacacacacacacac 1785
1753 atgagatacaaaaatactatacttgatcgaagaagaagaactgcttgtagatataatg 1812
1786 atgagatacaaaaatactatacttgatcgaagaagaagaactgcttgtagatataatg 1845
1813 gcttcaattatacacacacaaatccctgagagacatttggagacatgataataaacat 1872
1846 gcttcaattatacacacacaaatccctgagagacatttggagacatgataataaacat 1905
1873 ttttaattcaagtaacttcccccgtgtgaagttaactatggttgtgtgacaacttcaat 1932
1906 ttttaattcaagtaacttcccccgtgtgaagttaactatggttgtgtgacaacttcaat 1965
1933 ctatagaataatgaatggaagtggtggaatcttcaatttcaatgcttgagtggaacaaatg 1992
1966 ctatagaataatgaatggaagtggtggaatcttcaatttcaatgcttgagtggaacaaatg 2025
1993 ctatcaagagtgacaataaagttgaatgataatcaca 2032
2026 ctatcaagagtgacaataaagttgaatgataatcaca 2065
Db

RESULT 4
AAF97929 standard; cDNA; 2099 BP.
XX
XX AAF97929;
AC
XX
DT 01-JUN-2001 (first entry)
XX
DE Human secreted protein cDNA, SBO ID NO: 56.
XX
XX Human; secreted protein; immunomodulatory; antisclerotic;
KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KW neurotropic; anticonvulsant; antialzheimers; antiparkinsonian;

Dp	1400	ttaataacattgatttccttcgttttaatagatatttggaaataatgatcagaagaatgt	1459
Oy	1443	aaaacatttgaatagctcccggtatgtagaagaaaagtcgcgtgaatttatatgacaact	1502
Dp	1460	aaaacatttgaatagctccggtatgtagaagaaaagtcgcgtgaatttatatgacaact	1519
Oy	1503	acgaatgcttaactcttttaacacagcatagtgagaaatcatatttggctatgtatact	1562
Dp	1520	acgaatgcttaactcttttaacacagcatagtgagaaatcatatttggctatgtatact	1579
Oy	1563	atgacaatttgytaattgtcttaatttgatgttaatactctgaaacaagaagaagtt	1622
Dp	1580	atgacaatttgytaattgtcttaatttgatgttaatactctgaaacaagaagaagaagtt	1639
Oy	1623	tttaacttaagtagagcccttaaaatatgtatgtgcctatataacgcgttagtttggaaact	1682
Dp	1640	tttaacttaagtagagcccttaaaatatgtatgtgcctatataacgcgttagtttggaaact	1699
Oy	1683	gatatatgataaagaagagacgcgttttttaacccctcttcgaattttgtgaacct	1742
Dp	1700	gatatatgataaagaagagacgcgttttttttaacccctcttcgaattttgtgaacct	1759
Oy	1743	atggtcttaataatgatactaaataatactacatctgtatcttaagaagaacttagccttgyga	1802
Dp	1760	atggtcttaataatgatactaaataatactacatctgtatcttaagaagaacttagccttgyga	1819
Oy	1803	gatatatgatacttttcatataacacaaaatcccttggaggaaatttgaagcataga	1862
Dp	1820	gatatatgatacttttcatataacacaaaatcccttggaggaaatttgaagcataga	1879
Oy	1863	tataaacaattttatcttcagtaactttcccccgtgtgaagttaactatggtttgtgta	1922
Dp	1880	tataaacaattttatcttcagtaactttcccccgtgtgaagttaactatggtttgtgta	1939
Oy	1923	caactatctctataaataatgaatgagagtggtgtgaattcttaacttttgttggagt	1982
Dp	1940	caactatctctataaataatgaatgagagtggtgtgaattcttaacttttgttggagt	1999
Oy	1983	ggaccaatgtctatacagaatgtgacaaataagtaatgatgatctccaaa	2032
Dp	2000	ggaccaatgtctatacagaatgtgacaaataagtaatgatgatctccaaa	2049

RESULT	5
AAZ40848	
ID	AAZ40848 standard; DNA; 2010 BP.
AC	AAZ40848;
XX	
XX	
DT	18-JAN-2000 (first entry)
DE	Secreted protein EST coding sequence 33-77-4-E2-FL1.
XX	
XX	
KM	Secreted protein; fingerprint identification technique;
KM	chromosome mapping; human; hereditary disease; cancer;
KM	hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KM	autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
KM	renal injury; amino aciduria; hypoglycaemia; male rat infertility;
KM	hypertension; ss.
XX	
OS	Homo sapiens.
XX	
FN	W0940189-A2.
XX	
PD	12-AUG-1999.
XX	
PF	09-FEB-1999; 99WO-IB00282.
XX	
PR	09-FEB-1998; 98US-0074121.
PR	13-APR-1998; 98US-0081563.
PR	10-AUG-1998; 98US-0096116.
PR	04-FEB-1998; 98US-0099273.
XX	

PA	(GEST) GENSET.
PX	
PI	Bouquelieret L., Duclert A., Dumas Milne Edwards J;
DR	WPI: 1999-600966/51.
XX	P-PsDB; AAy59720.
PR	Extended cDNAs useful for expressing secreted proteins and to obtain specific antibodies -
PS	Claim 1: Page 227-229; 244pp: English.
CC	This sequence encodes a human secreted protein of the invention. The extended cDNAs (or genomic DNAs obtainable from them) may be used to prepare PCR primers and probes. These are useful for forensic matching or positive identification by DNA sequencing. They may also be used in alternative fingerprint identification techniques. Antibodies against the CC proteins encoded by the extended cDNAs are useful in identification of tissue types or cell species, as well as identifying tissue specific soluble proteins. The sequences can be used for chromosome mapping and identification of genes associated with hereditary diseases or drug response. signal sequences from the cDNAs can be used in construction of secretion vectors. Other sequences derived from the extended cDNAs can be used to clone upstream genomic DNA sequences including promoters. This is CC in turn useful for identifying proteins that interact with promoter CC sequences. Some of the proteins may be useful in diagnosing and treating several disorders including, but not limited to: cancer, hyperlipidaemia, cardiovascular and neurodegenerative disorders, autoimmune diseases, and rheumatic diseases, embryogenic disorders, hypertension, renal injury, amino acidurias, hypoglycaemia, male rat infertility and myopathies.
SQ	Sequence 2010 BP; 565 A; 360 C; 431 G; 642 T; 12 other:
Query Match	94.2%; Score 1914.8; DB 20; Length 2010;
Best Local Similarity	98.2%; Pred. No. 0;
Matches 1966; Conservative	10; Mismatches 22; Indels 4; Gaps 4;
QY	34 tgcgccgctgtcccgccggccgctcgttgcgcccgagtgctgcggccgcccgcgcgc 93 :::
Db	2 tgccgccgcttcgcccgccgccccgttcgtgcgcccgacgtgtgcgcgcgcgcckc 61
QY	94 accatggcctgttttgtgcacctgcgggtttgttgccgggtgttttgggaagaactt 153
Db	62 accaatggcctgttgcgtgcgcctgcgggttgcgggttgcgggttgcgggttgcgggtt 121
QY	154 agtatataaacatcacccaggctcgttgccttcgcgaatvgaaattgacctataccaga 213
Db	122 agtatataaacatcacccaggctcgttgccttcgcgaatvgaaattgacctataccaga 181
QY	214 gggcgagatccccaagctgagctgatgtccatggagctctctcgttaaaaagaacccttct 273
Db	182 gagcgagatccccaagctgagctgatgtccatggagctctctcgttaaaaagaacccttct 241
QY	274 tggcccaagactcgcagtggtgaactcgtltaactgcgtcctccggccatccgcatacgtgtga 333
Db	242 tggcccaagactcgcagtggtgaactcgtltaactgcgtcctccggccatccgcatacgtgtga 301
QY	334 gtgaaggagtgacaacaactgagctcaccaccaagcagtgatattgtaacccttttag 393
Db	302 gtgaaggagtgacaacaactgagctcaccaccaagcagtgatattgtaacccttttag 361
QY	394 aatggagtlctcctttaagtcttgagacgtgttgaatatcatcactcactatttttcag 453
Db	362 aatggagtlctcctttaagtcttgagacgtgttgaatatcatcactcactatttttcag 421
QY	454 gaaactcgttttttcgaattggtcctcccgatgaagaagaagtgatatacgtgaaggag 513
Db	422 gaaactcgttttttcgaattggtcctcccgatgaagaagaagtgatatacgtgaaggag 481
QY	514 gaaactcagtglttgaagaccttcaagcacttcggccgcgcgcactcgtlaactgcgcgttt 573
Db	482 gomaactcagtglttgaagaccttcaagcacttcggccgcgcgcactcgtlaactgcgcgttt 540

QY 574 caagaaactcgtttccagttcaactccccccaattc-lcgtagtaggaacaatgaagt 632
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 541 caagaaactcgtttccagttcaactccccccaattc-lcgtagtaggaacaatgaagt 600
 QY 633 tgaactcgtttcttcttctgaactgaactgaactgaactgaactgaactgaactgaact 692
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 601 tgaactcgtttcttcttctgaactgaactgaactgaactgaactgaactgaactgaact 660
 QY 693 tgaactcgtttcttcttctgaactgaactgaactgaactgaactgaactgaactgaact 752
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 661 tgaactcgtttcttcttctgaactgaactgaactgaactgaactgaactgaactgaact 720
 QY 753 ggaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 812
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 721 ggaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 780
 QY 813 ccttctgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 872
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 781 ccttctgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 840
 QY 873 agtgaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 932
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 841 agtgaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 900
 QY 933 taccctgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 992
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 901 taccctgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 960
 QY 993 taattgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1052
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 961 taattgaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1020
 QY 1053 ggcctgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1111
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1021 ggcctgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1080
 QY 1112 ttatagaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1170
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1081 ttatagaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1140
 QY 1171 gaaagaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1230
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1141 kaaagaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1200
 QY 1231 agatagaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1290
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1201 agatagaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1260
 QY 1291 ctgtgaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1350
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1261 ctgtgaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1320
 QY 1351 ccatagaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1410
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1321 ccatagaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1380
 QY 1411 atgaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1470
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1381 atgaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1440
 QY 1471 aaaaagaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1530
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1441 aaaaagaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1500
 QY 1531 aggtgaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1590
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1501 aggtgaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1560
 QY 1591 atgtaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1650
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1561 atgtaactgaactgaactgaactgaactgaactgaactgaactgaactgaactgaact 1620

QY 1651 atgtctataatactcgttagtttgggaactgatactgataagaagagagcgtgtt 1710
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1621 atgtctataatactcgttagtttgggaactgatactgataagaagagagcgtgtt 1680
 QY 1711 tttaacctctctcgaactgttctgacctacacagggctaaataatgaactaaataact 1770
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1681 tttaacctctctcgaactgttctgacctacacagggctaaataatgaactaaataact 1740
 QY 1771 acatgactagaagaagaactgacctgttgagatagatgactgtttcatataaacac 1830
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1741 acatgactagaagaagaactgacctgttgagatagatgactgtttcatataaacac 1800
 QY 1831 aaaaacctcgaaggaactgttgagacataataaataatattatgaact 1890
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1801 aaaaacctcgaaggaactgttgagacataataaataatattatgaact 1860
 QY 1891 tccccctgtgaactgactgttctggtacacactcattcctataaataatgaact 1950
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1861 tccccctgtgaactgactgttctggtacacactcattcctataaataatgaact 1920
 QY 1951 aagtgggtgaacttactatttattgttgagtgagacacatgctatacaagatgacaaat 2010
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1921 aagtgggtgaacttactatttattgttgagtgagacacatgctatacaagatgacaaat 1980
 QY 2011 aagtgaactgacttccaaa 2032
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1981 aagtgaactgacttccaaa 2002

RESULT 6
 AAV49561
 ID AAV49561 standard; cDNA to mRNA; 1050 BP.
 AC
 XX AAV49561:
 DT 21-OCT-1998 (first entry)
 XX
 DE Human epidermal carcinoma cell line KB clone HP01293 cDNA #2.
 KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;
 KW haemostasis; activin; inhibitor; chemotactic; chemokinetic; receptor;
 KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO9821328-A2.
 XX
 PD 22-MAY-1998.
 XX
 PE 07-NOV-1997; 97MO-JP04056.
 XX
 PR 13-NOV-1996; 96JP-0301429.
 XX
 PA (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 XX
 PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
 DR WPI: 1998-297932/26.
 DR P-PSDB; AAM64539.
 XX
 PT Human protein having transmembrane domain - useful for, e.g.
 PS research and nutrition
 XX
 PS Claim 3: Page 121-122; 205pp; English.
 CC AAV49550-V49599 are cDNA sequences which encode human proteins containing
 CC a transmembrane domain. These proteins can be used for, e.g. research
 CC and nutrition, and may have cytokine and cell
 CC proliferation/differentiation, immune stimulating/suppressing,
 CC haemopoiesis regulating, tissue growth, activin/inhibin,
 CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,

CC anti-inflammatory or tumour inhibition activity.

XX Sequence 1050 BP; 261 A; 218 C; 253 G; 318 T; 0 other:

Query Match 51.6%; Score 1050; DB 19; Length 1050;

Best Local Similarity 100.0%; Pred. No. 1.8e-253;

Matches 1050; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 97 atggcgtgttggcgtgccccgcgttggtggcggtgttttggggaagagtaagt 156
   |||||||
DB 1 atggcgtgttggcgtgccccgcgttggtggcggtgttttggggaagagtaagt 156
   |||||||
QY 157 alattaataccagaggtctgtgttcttcgaalaggaattgacctataccagagag 216
   |||||||
DB 61 alattaataccagaggtctgtgttcttcgaalaggaattgacctataccagagag 216
   |||||||
QY 217 cggatcccaagagtggtgctgcatgttccatgggtcttcctctgtgaagaagaccttcttg 276
   |||||||
DB 121 cggatcccaagagtggtgctgcatgttccatgggtcttcctctgtgaagaagaccttcttg 276
   |||||||
QY 277 ccagagctcgagtggtgtaacctgttcaatcgtctcgtggctacccgtcagtgatgtg 336
   |||||||
DB 181 ccagagctcgagtggtgtaacctgttcaatcgtctcgtggctacccgtcagtgatgtg 336
   |||||||
QY 337 aagagagtgaaacaaactggtcttaaccccaagcagtgatcttcgtacaccttgagaat 396
   |||||||
DB 241 aagagagtgaaacaaactggtcttaaccccaagcagtgatcttcgtacaccttgagaat 396
   |||||||
QY 397 gcagttccctttagctcttgacaagtgtgtcaaatccatccatcttcttgagaa 456
   |||||||
DB 301 gcagttccctttagctcttgacaagtgtgtcaaatccatccatcttcttgagaa 456
   |||||||
QY 457 acccgtgtgtttgagtggtggtcccaagtgagaagaatgtatgttgggaagaa 516
   |||||||
DB 361 acccgtgtgtttgagtggtggtcccaagtgagaagaatgtatgttgggaagaa 516
   |||||||
QY 517 aactcagtggttgagaagaccttcaagcttcagctgcagactccgtatccgttcttcaa 576
   |||||||
DB 421 aactcagtggttgagaagaccttcaagcttcagctgcagactccgtatccgttcttcaa 576
   |||||||
QY 577 gaaactcgtgttcccaatcttaaccccccaatctctctgagtgagaagaatgtgac 636
   |||||||
DB 481 gaaactcgtgttcccaatcttaaccccccaatctctctgagtgagaagaatgtgac 636
   |||||||
QY 637 ctgctcttcttcttgacatgcaagtgtacatgatatcttcaagcttgcgtccgtcat 696
   |||||||
DB 541 ctgctcttcttcttgacatgcaagtgtacatgatatcttcaagcttgcgtccgtcat 696
   |||||||
QY 697 aagcactagccaagatctctctctgattatatactctgagctgagctgttgat 756
   |||||||
DB 601 aagcactagccaagatctctctctgattatatactctgagctgagctgttgat 756
   |||||||
QY 757 gaaatgggaagcgttcttggaagactctgaacaatctgaatcttccaatctct 816
   |||||||
DB 661 gaaatgggaagcgttcttggaagactctgaacaatctgaatcttccaatctct 816
   |||||||
QY 817 gttgaagctctgcaaaagtgtgacatgacatgacagcttcttaagtgggaatgacgtg 876
   |||||||
DB 721 gttgaagctctgcaaaagtgtgacatgacatgacagcttcttaagtgggaatgacgtg 876
   |||||||
QY 877 gtagagttagtcactgcaagcatgtacacccctccctcttaaggaagaagacatc 936
   |||||||
DB 781 gtagagttagtcactgcaagcatgtacacccctccctcttaaggaagaagacatc 936
   |||||||
QY 937 ctggaagcaaaacgaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 996
   |||||||
DB 841 ctggaagcaaaacgaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 996
   |||||||
QY 997 ttggaatctcgtgtgttccaacatgtaattgtaatagtcgttgccttgagct 1056
   |||||||
DB 901 ttggaatctcgtgtgttccaacatgtaattgtaatagtcgttgccttgagct 960

```

RESULT 7

AAZ17313 standard; cDNA: 793 BP.

AAZ17313;

12-OCT-1999 (first entry)

Human gene expression product cDNA sequence SEQ ID NO:4786.

Human; gene: gene expression product; diagnosis; therapy; probe;

detection; mapping; tissue typing; profiling; forensic; cancer;

genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

Homo sapiens.

W0938972-A2.

05-AUG-1999.

28-JAN-1999; 99WO-US01619.

03-APR-1998; 98US-0080666.

28-JAN-1998; 98US-0072910.

24-FEB-1998; 98US-0075954.

31-MAR-1998; 98US-0080114.

03-APR-1998; 98US-0080515.

(CHIR) CHIRON CORP.

(HYSE-) HYSECO INC.

Claim 1, Page 2275-2276; 2479p; English.

The present invention describes a library of human polynucleotides

comprising the sequences given in AAZ12532 to AAZ17779. Also described is

a method of detecting differentially expressed genes correlated with the

cancerous state of a mammalian cell, comprising detecting at least one

differentially expressed gene product in a test sample from a cell

suspected of being cancerous, where the gene product is encoded by one

of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The

polynucleotides can be used as a source of primers and probes, which can

be used for a variety of purposes, e.g. detection of expression levels,

mapping, tissue typing or profiling, forensics, genetic analysis and

detection of polymorphisms. Polypeptides encoded by the polynucleotides

can be used for raising antibodies for experimental, diagnostic and

CC peptide analogues and antagonists.

XX Sequence 793 BP; 260 A; 108 C; 143 G; 267 T; 15 other;

Query Match 31.4%; Score 639; DB 20; Length 793;

Best Local Similarity 97.9%; Pred. No. 1.5e-150;

Matches 697; Conservative 0; Mismatches 10; Indels 5; Gaps 5;

OY 1311 gaattatagttatgtaacccactggtatagattatcattatggttaatt 1370
 |||||||
 DB 58 gaattatagttatgtaacccactggtatagattatcattatggttaatt 117
 |||||||
 OY 1371 atgatatagcatttaataacattgattcattctctgtttaatgaatttggaatagac 1430
 |||||||
 DB 118 atgatatagcatttaataacattgattcattctctgtttaatgaatttggaatagac 177
 |||||||
 OY 1431 tgaagaatgtaaacattgtaagtcgtgtgtatggaaaaagtcgactgaattta 1490
 |||||||
 DB 178 tgaagaatgtaaacattgtaagtcgtgtgtatggaaaaagtcgactgaattta 237
 |||||||
 OY 1491 ttagaacaacttaagaaatgcttaactctttacacagcagataggtgaatcatattgg 1550
 |||||||
 DB 238 ttagaacaacttaagaaatgcttaactctttacacagcagataggtgaatcatattgg 296
 |||||||
 OY 1551 ctattgttactatgaacaatttgaataatgcttaattgtaatgaataactcgaaca 1610
 |||||||
 DB 297 ctattgttactatgaacaatttgaataatgcttaattgtaatgaataactcgaaca 356
 |||||||
 OY 1611 agagaaagtttttaacttagagtagccccaataatagatgtgtctataatagcgtt 1670
 |||||||
 DB 357 agagaaagtttttaacttagagtagccccaataatagatgtgtctataatagcgtt 416
 |||||||
 OY 1671 agtttgaactgtatctggtatgaacagagacgctgttttaaccccttcgcgaat 1730
 |||||||
 DB 417 agtttgaactgtatctggtatgaacagagacgctgttttaaccccttcgcgaat 476
 |||||||
 OY 1731 ttgttgaactcatggtgcttaataatgatacctaanaactacatgacttaagaagaac 1790
 |||||||
 DB 477 ttgttgaactcatggtgcttaataatgatacctaanaactacatgacttaagaagaac 536
 |||||||
 OY 1791 tagccttgagagatataagatgctttcattatcacacaaaatcccttaagggaact 1850
 |||||||
 DB 537 tagccttgagagatataagatgctttcattatcacacaaaatcccttaagggaact 596
 |||||||
 OY 1851 ctgaagcatgataataaacattttatttcagtaacctttcccccgtgt-aaqtact 1909
 |||||||
 DB 597 ctgaagcatgataataaacattttatttcagtaacctttcccccgtgtlaaagttact 656
 |||||||
 OY 1910 atggtt-ctgtgtaacactcattcattatagataatgaat-aggaaatggtgaattctac 1967
 |||||||
 DB 657 atggttgggggtacacattcattcattatagataatgaatagggaagtggtgaattctac 716
 |||||||
 OY 1968 tttttat-gttgaggtgaccaaagtctataagaagtgaacaattta 2018
 |||||||
 DB 717 tttttatggttgggggtgagcaactgctatacaagagacaatnaagttta 768
 |||||||
 RESULT 8
 AAX98768
 ID AAX98768 standard; cDNA; 793 BP.
 AC AAX98768;
 XX
 XX 24-SEP-1999 (first entry)
 DE Human validated cancer cell derived cDNA #110.
 XX
 XX Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
 KM Integral membrane protein; aspartyl protease; GATA family; wnt family;
 KM transcription factor; G-protein alpha subunit; protein phosphatase;
 KM phospholipase binding protein; diacylglycerol binding protein; trypsin;
 KM protein kinase; tyrosine phosphatase; developmental signalling protein;

KM Wnt/sp5/MMP domain; therapy; forensic; genetic mapping; diagnostic;
 KM detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
 KM Wilms' tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
 KM leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
 KM prostate; ss.

XX Homo sapiens.

XX W09933982-A2.

PN 08-JUL-1999.

XX 22-DEC-1998; 98WO-US27610.

XX 21-DEC-1998; 98US-0217471.

PR 23-DEC-1997; 97US-0068755.

PR 03-APR-1998; 98US-0080664.

PR 21-OCT-1998; 98US-0105234.

PR 27-OCT-1998; 98US-0105877.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Chkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

PI Jones LM, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-430243/36.

XX New isolated human polynucleotides

XX Claim 1: Page 457-458; 591pp; English.

XX This invention describes novel isolated human polynucleotides obtained

CC by screening for differential expression in colon cancer, breast cancer

CC and lung cancer cell lines. The polynucleotides of the invention are

CC represented in AAX98275-X9918 and encode polypeptides of protein

CC families selected from 4 transmembrane segments integral membrane

CC proteins, 7 transmembrane receptors, ATPases associated with various

CC cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of

CC transcription factors, G-protein alpha subunit, phospholipases or

CC diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,

CC protein tyrosine phosphatase, trypsin, wnt family of developmental

CC signalling proteins and Wnt/sp5/MMP domain containing proteins. The

CC encoded polypeptides also have a functional domain selected from Ank

CC repeat, basic region plus leucine zipper transcription factors,

CC bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger

CC (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease

CC domain. The polynucleotides encode polypeptides with similarity to known

CC protein families and are predicted to have similar properties. The novel

CC polynucleotides can be used to develop products for use as therapeutic

agents and in forensics, genetic analysis, mapping and diagnostic

CC applications. In particular, the product can be used for the detection

CC and management of cancers. They can be used for treating e.g. cervical

CC cancers, melanomas, colorectal adenocarcinomas, Wilms' tumour, sarcomas,

CC retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic

CC myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and

CC myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydic

CC hereditary ectodermal dysplasia, congenital alveolar dysplasia,

CC epithelial dysplasia of the cervix, fibrous dysplasia of bone, and

CC mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,

CC prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of

CC the skin.

XX

XX Sequence 793 BP; 260 A; 108 C; 143 G; 267 T; 15 other;

Query Match 31.4%; Score 639; DB 20; Length 793;

Best Local Similarity 97.9%; Pred. No. 1.5e-150;

Matches 697; Conservative 0; Mismatches 10; Indels 5; Gaps 5;


```

OY 1311 gaattatagcttgacgtgaatccacgtgtgtatgaattccataatgcttcaatt 1370
DB 58 gaattatagcttgacgtgaatccacgtgtgtatgaattccataatgcttcaatt 117
OY 1371 atgataagcaattataacattgatttccctgcttcaattgaattggaataatcac 1430
DB 118 atgataagcaattataacattgatttccctgcttcaattgaattggaataatcac 177
OY 1431 tgaagaagaattgaataatcattagaatgctgtgtatggaanaaagtcacgtcaatta 1490
DB 178 tgaagaagaattgaataatcattagaatgctgtgtatggaanaaagtcacgtcaatta 237
OY 1491 ttgagaaactgaagatgcttcaacttcttaccacagataggtgaanaatcatttggg 1550
DB 238 ttanacaactgaagatgcttcaacttcttaccacagataggtg-aaatcaatttggg 296
OY 1551 ctattgtatactatgaacaatttgttaattgtccttaatttgaatgaataacttgaaca 1610
DB 297 ctattgtatactatgaacaatttgttaattgtccttaatttgaatgaataacttgaaca 356
OY 1611 agagaaagttttaaacttagagtagccctaaatatgatgtgtattataatgcgtt 1670
DB 357 agagaaagttttaaacttagagtagccctaaatatgatgtgtattataatgcgtt 416
OY 1671 agtttgaactgtatctgtgtgaacagagagacgttctttaaacttcttgcgaagt 1730
DB 417 agtttgaactgtatctgtgtgaacagagagacgttctttaaacttcttgcgaagt 476
OY 1731 ttgttgaactcatggtgcatatagatgaataaataactatcatgattgaagaagaac 1790
DB 477 ttgttgaactcatggtgcatatagatgaataaataactatcatgattgaagaagaac 536
OY 1791 taaccttgaagatgaatgaatgcttcttcatatatacacaanaaactccgtgaagaact 1850
DB 537 taaccttgaagatgaatgaatgcttcttcatatatacacaanaaactccgtgaagaact 596
OY 1851 ttgagagcatgaataaacaattttattcagtaaatcttcccccgtgtc-aagttact 1909
DB 597 ttgagagcatgaataaacaattttattcagtaaatcttcccccgtgtc-aagttact 656
OY 1910 atggtttt-gtggtaacttcttcatatagaataatgaat-ggaagtggtggaattctac 1967
DB 657 atggttttgggggtacaaacttcttcatatagaataatgaatggtggaattctac 716
OY 1968 tttttat-gtctgagtggaacaatgtctatacaagatggaacaataaagttaa 2018
DB 717 tttttatggttgggtggaacaatgtctatacaagatggaacaataaagttaa 768

```

RESULT 9

AAV87120 standard; cDNA; 577 BP.

AAV87120;

27-APR-1999 (first entry)

EST clone B087.

Expressed sequence tag; secreted protein; haematopoiesis regulator;
tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
chemokinesis; chemokinesis; haemostasis; gene therapy; thrombolysis;
receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

Homo sapiens.

M09845435-A2.

15-OCT-1998.

10-APR-1998; 98MO-US06954.

10-APR-1997; 97US-0835913.

```

XX (GEMV ) GENETICS INST INC.
PA Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 1999-070076/06.
XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX
XX Claim 1: Page 459-460; 633pp; English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokine activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
XX
SO Sequence 577 BP; 195 A; 84 C; 102 G; 196 T; 0 other;

```

Query Match

Best Local Similarity 99.5%; Pred. No. 1,3e-127;

Matches 549; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY 1298 gttttctagagtgatattagatgacgtgaatccacgtgtgtatgaattccat 1357
DB 22 gttttctagagtgatattagatgacgtgaatccacgtgtgtatgaattccat 81
OY 1358 atgcttgaattatgatataagcatttaataatgatttcaattctgttaagaatt 1417
DB 82 atgcttgaattatgatataagcatttaataatgatttcaattctgttaagaatt 141
OY 1418 ttgaaatctgacgtgaagaagaatgtgaaacattagatagctgtgtatggaanaag 1477
DB 142 ttgaaatctgacgtgaagaagaatgtgaaacattagatagctgtgtatggaanaag 201
OY 1478 ttgactgaattatttagacaacatttagaatgttaacttcttacaagatagtgaa 1537
DB 202 ttgactgaattatttagacaacatttagaatgttaacttcttacaagatagtgaa 261
OY 1538 aatcatattgggtatgtatgaactatgaacaattgttaagtcttattatgaataa 1597
DB 262 aatcatattgggtatgtatgaactatgaacaattgttaagtcttattatgaataa 321
OY 1598 taactctgaacaagaagaagttttaaacttagagtagcctaaatatgagtgct 1657
DB 322 taactctgaacaagaagaagttttaaacttagagtagcctaaatatgagtgct 381
OY 1658 tataaactcgcttattgttgaacgtatctgtgaacagagagacgtgtttttaaac 1717
DB 382 tataaactcgcttattgttgaacgtatctgtgaacagagagacgtgtttttaaac 441
OY 1718 ctctctcgcaagttgttgacctacatgagcctaataatgataacataacatgaa 1777
DB 442 ctctctcgcaagttgttgacctacatgagcctaataatgataacataacatgaa 501
OY 1778 tctaagaagaactagccttggagatataagatgcttcatatatacacaanaaact 1837
DB 502 tctaagaagaactagccttggagatataagatgcttcatatatacacaanaaact 561
OY 1838 cctgagggagacat 1849

```

Db 562 cctgagtgact 573

RESULT 10
AAZ15750
AAZ15750 standard; cDNA; 813 BP.

AC AAZ15750;
XX
DT 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:3219.

XX Human; gene: gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

PN WO938972-A2.

PD 05-AUG-1999.

XX 28-JAN-1999; 99WC-US01619.

XX 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Ctkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

PI Jones WJ, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Suduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

DR Novel human genes and their expression products which are

PT differentially expressed in different cell types

XX Claim 1, Page 1545; 2479pp; English.

The present invention describes a library of human polynucleotides comprising the sequences given in AAZ1532 to AAZ1779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AAZ1532 to AAZ1779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

XX Sequence 813 BP; 251 A; 123 C; 143 G; 267 T; 29 other;

Query Match 26.9%; Score 547; DB 20; Length 813;
Best Local Similarity 92.3%; Pred. No. 1,7e-127;
Matches 651; Conservative 0; Mismatches 38; Indels 16; Gaps 7;

OY 1311 gaattatagatgacgagacccactggtgatagatccataatagcttgatatt 1370
DB 93 gaattatagatgacgagacccactggtgatagatccataatagcttgatatt 152
OY 1371 atgatagccattatatacatgattcattcctgtttaatgaattggaaatgac 1430
DB 153 atgatagccattatatacatgattcattcctgtttaatgaattggaaatgac 212
OY 1431 tgaagaagatgaaacatttagaagctcgtgttaagaaagatgacgattta 1490
DB 213 tgaagaagatgaaacatttagaagctcgtgttaagaaagatgacgattta 272
OY 1491 ttagaacaacttcgaatgcttaactcctttacacagcagatgagaaatcatatttg 1550
DB 273 ttagaacaacttcgaatgcttaactcctttacacagcagatgagaaatcatatttg 332
OY 1551 ctattgtactatgaaacatttgtaagcttcaatttgatgtaaatcaacttgaaca 1610
DB 333 ctattgtactatgaaacatttgtaagcttcaatttgatgtaaatcaacttgaaca 392
OY 1611 agagaagaaggttttaacttaagagtagccctaaatgagatgagctatataatgctt 1670
DB 393 agagaagaaggttttaacttaagagtagccctaaatgagatgagctatataatgctt 452
OY 1671 agttttgaaactgatactgaagtaacagagacagctgtttttaaaccctctgcaagt 1730
DB 453 agttttgaaactgatactgaagtaacagagacagctgtttttaaaccctctgcaagt 511
OY 1731 ttgtgacactaactgagcctaatalatgatactaaataactatgattcctaaagaagac 1790
DB 512 ttgtgacactaactgagcctaatalatgatactaaataactatgattcctaaagaagac 571
OY 1791 tagcctgtgagatataatagatgctttcattataca-caaaaaccctggagacat 1849
DB 572 tagcctgtgagatataatagatgctttcattataca-caaaaaccctggagacat 631
OY 1850 ttgagagatgaataataaacttttattt---cagtaactttccctcgtgt-aagt 1905
DB 632 tttnagatgaatataataaacttttatttcaagtaactttccctcgtgtgaagt 691
OY 1906 tactatggttctgag---tacaactcattatagaataatagtgaa-----gtgggt 1958
DB 692 tactatggttctgag---tacaactcattatagaataatagtgaa-----gtgggt 751
OY 1959 gaattcactttatgttgag--tgacaaatgctatcaag 2000
DB 752 aaattcactttatgttgagtggtggcccaatgctatcaag 796

RESULT 11

AAZ15759
ID AAZ15759 standard; cDNA; 813 BP.

XX AAZ15759;

AC AAZ15759;

XX 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:3228.

XX Human; gene: gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

PN WO938972-A2.

PD 05-AUG-1999.

CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretion
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantify the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
CC (ELISA)). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.

SO Sequence 588 BP; 202 A; 114 C; 74 G; 193 T; 5 other;

Query Match 25.2%; Score 512.6; DB 22; Length 588;
Best Local Similarity 97.2%; Pred. No. 6.2e-119;

Matches 551; Conservative 0; Mismatches 13; Indels 3; Gaps 3;

OY 1463 tctatggaagaaagtcgactgaattatgacaaactgacgaatgctactcttla 1522
DB 565 tcttttgggagaaagtcgactgaattatgacaaactgacgaatgctactcttla 508
OY 1533 cacaagcatggtgaaagaa-tcattatggtgctatgctactgacaaattgttaatt 1581
DB 507 CACAGCATGAGTGGGAAATCATATTTGGGCTATGTTACTATGAACAAATTTGTAATGT 448
OY 1582 cttacttgaatgaataaactcgaacagaagaaggttctactgaagagccct 1641
DB 447 CTTAATTGATTAATTAATCTGGAACAGAGNAAGGTTTAACTTAAGATAGCCCT 368
OY 1642 aaaaatggaatgctactatgaatcagctagtttgaactgtatcgaagaaga 1701
DB 387 AAAATATGATGTGGTATATATCCCTTATGTTGGAACCTGATATGAGACAGAGA 328
OY 1702 cagcttcttctaaacctctcgaagtttgaactcgaatgaggtcctatgact 1761
DB 327 CAACTGTTTTTAAACCTCTTGCAGAGTTTGTGACCTACATGAGGCTAATATGATAT 268
OY 1762 aaaaatggaatgctactatgaatcagctagtttgaactgtatcgaagaaga 1821
DB 267 AAAATATGATGTGGTATATATCCCTTATGTTGGAACCTGATATGAGACAGAGA 208
OY 1822 tatacacaagaaatccctgaggaacatttgaagcattgaaataaacttctatc 1881
DB 207 TATACACACAAAATCCCTGAGGACATTTGAGGCAATGATTAACATTTTATTTC 148
OY 1882 agtaacttccctcctgtgtaagtactactggttgtgtaacttactatagaat 1941
DB 147 AGTAACTTTTCCCTCTGTATGATGATGATGATGATGATGATGATGATGATGAT 88
OY 1942 atcaagtgaagtggaatctacttctatgcttgaagtggaacaaatgctatcaga 2001
DB 87 ATTAAGTGAAGTGGTGAATTTTACTTTTATGTTGAGAGTGAACCAATGTTATCAGA 28
OY 2002 gtaacaaataaagtatgattcc 2028
DB 27 GTGACAAATTAATGATGATTC 1

RESULT 13

AAF93970 ID AAF93970 standard; DNA; 560 BP.

XX AAF93970;

XX 23-MAY-2001 (first entry)

XX

DE Primer specific for DNA encoding secretory/membrane protein SEQ ID 404.

XX Human; secretory protein; membrane protein; vaccine; gene therapy;

KW Rheumatoid arthritis; diabetes; PCR primer; ss.

XX Synthetic.

OS EPI067182-A2.

XX 10-JAN-2001.

XX 07-JUL-2000; 2000EP-0114090.

XX 08-JUL-1999; 99JP-0194179.

PR 11-JAN-2000; 2000JP-0118775.

PR 02-MAY-2000; 2000JP-0183766.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

XX WPI; 2001-093989/11.

XX Claim 4; SEQ ID 404; 609pp + CD ROM; English.

XX This invention relates to nucleic acid sequences AAF93744 - AAF93916

CC which encode human secretory or membrane proteins represented by

CC AAB88317 - AAB88419. Included in the invention are primers

CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the

CC cDNA sequences of the invention. The invention also includes methods for

CC the production of antibodies directed against the proteins, and cDNA

CC sequences, which can be used in vaccines. The polynucleotide sequences

CC can be used in gene therapy. The polynucleotide sequences and the

CC proteins they encode may be used in the prevention, treatment and

CC diagnosis of diseases associated with inappropriate secretory

CC protein/membrane protein expression. The nucleic acids and complementary

CC sequences may also be used as DNA probes in diagnostic assays

CC (e.g. polymerase chain reactions (PCR)) to detect and quantify the

CC presence of similar nucleic acid sequences in samples. They may also be

CC used to study the expression and function of secretory proteins/membrane

CC polypeptides and their role in metabolism. The polypeptides may be used

CC as antigens in the production of antibodies against them and in assays to

CC identify modulators (agonists and antagonists) of expression and

CC activity. The antibodies and antagonists may also be used as therapeutic

CC agents to down regulate expression and activity. The antibodies may also

CC be used as diagnostic agents for detecting the presence of the

CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay

CC (ELISA)). Examples of diseases which may be treated include rheumatoid

CC arthritis and diabetes.

XX Sequence 560 BP; 104 A; 145 C; 163 G; 142 T; 6 other;

Query Match 23.9%; Score 486.8; DB 22; Length 560;

Best Local Similarity 98.6%; Pred. No. 1.8e-112;

Matches 488; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 gactcgaagcgcgtacactctcctcaagctgagctgtgcccgtgtccgcgcgcgttc 60
DB 66 gactcgaagcgcgtacactctcctcaagctgagctgtgcccgtgtccgcgcgcgttc 125
OY 61 cgtgcgcgcgcgcgtcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
DB 126 cgtgcgcgcgcgcgcgtcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 185
OY 121 gctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
DB 186 gctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 245
OY 181 gtttcgcaatggaattgactatcaccagagagcgatccaccagagtgctgctatg 240

```

Db 246 gtcttcgaaatggaatcgtccataccagagagcgtaccacagctgctgcatg 305
Oy 241 tcatgaggtctctctgtgaaagaagaccttcttgcgcagactcagtggaactg 300
Db 306 tcatgaggtctctctgtgaaagaagaccttcttgcgcagactcagtggaactg 365
Oy 301 tttatcgtccctcgggctccacgtcgtggtgagtggaaggaatgaaacactgacct 360
Db 366 tttatcgtccctcgggctccacgtcgtggtgagtggaaggaatgaaacactgacct 425
Oy 361 ccccgagagtgatctatctgtacaccttggagaatgcagttcctttagtctgacagt 420
Db 426 ccccgagagtgatctatctgtacaccttggagaatgcagttcctttagtctgacagt 485
Oy 421 gttgcaatcattccattcttcttcttggagaatcctctgtgttttgacgttgc 480
Db 486 gttgcaatcattccattcttcttcttggagaatcctctgtgttttgacgttgc 545
Oy 481 cccagtgaggaaga 495
Db 546 cccantgaggaaga 560

RESULT 14
AA215793
ID AA215793 standard; cDNA; 858 BP.
AC AA215793;
XX
XX 12-OCM-1999 (first entry)
DE Human gene expression product cDNA sequence SEQ ID NO:3262.
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
OS Homo sapiens.
XX
XX W09938972-A2.
PN 05-AUG-1999.
XX
XX 28-JAN-1999; 99MO-US01619.
PE
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crivenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Gliese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Studduth-Klinger J, Williams LT;
XX
XX WPI; 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1; Page 1565; 2479p; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AA21532 to AA21779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell

```

```

CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AA21532 to AA21779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
XX Sequence 858 BP; 205 A; 186 C; 189 G; 225 T; 53 other;
SQ

```

```

Query Match 19.0%; Score 386.2; DB 20; Length 858;
Best Local Similarity 92.4%; Pred. No. 3,4e-87;
Matches 465; Conservative 0; Mismatches 30; Indels 8; Gaps 6;

```

```

Oy 189 aaatgaaatgagctataccagagagcgtaccagacgtgctgcatgtccatgag 248
Db 59 aaacgtctgggctataccagagagcgtaccagacgtgctgcatgtccatgag 118
Oy 249 ctctctgtgaaagaagaccttcttgcgcagactcagtggtgaacctgtttacg 308
Db 119 ctctctgtgaaagaagaccttcttgcgcagactcagtggtgaacctgtttacg 178
Oy 309 tccctgggtaccgctcat-ggtgagtgaaaggagtgaaacactgctctaccacc 367
Db 179 nccmcygctaccgctcatggtgagtgaaaggagtgaaacactgctctaccacc 238
Oy 368 gcatgctatcttgacaccttgagaaatgcagcttctttagcttgaacagttgcaa 427
Db 239 gcagntcatcttgacaccttgagaaatgcagcttctttagcttgaacagttgcaa 298
Oy 428 attccattacaccttattcttctgagaaactcctgtgttctgcaagttgctccag 487
Db 299 attccattacaccttattcttctgagaaactcctgtgttctgcaagttgctccag 358
Oy 488 agaaagagtgatagtgtaggaagcaactcagtggttgaagaccttgaacct 547
Db 359 agaaagagtgatagtgtaggaagcaactcagtggttgaagaccttgaacct 418
Oy 548 t---gcgcagctccgtaatgcctgttccaagaaactcgttctcag-ttcaacctcc 603
Db 419 ttgcgcagctccgtaatgcctgttccaagaaactcgttctcag-ttcaacctcc 478
Oy 604 ctcaattctcgtagtaggaacatg-aagtgaacctgtc-ttcttcttgaacctgaag 661
Db 479 ctcaattctcgtagtaggaacatg-aagtgaacctgtc-ttcttcttgaacctgaag 538
Oy 662 tgcctac-atgatattcaagctt 683
Db 539 tgcctacatgatattcaagctt 561

```

```

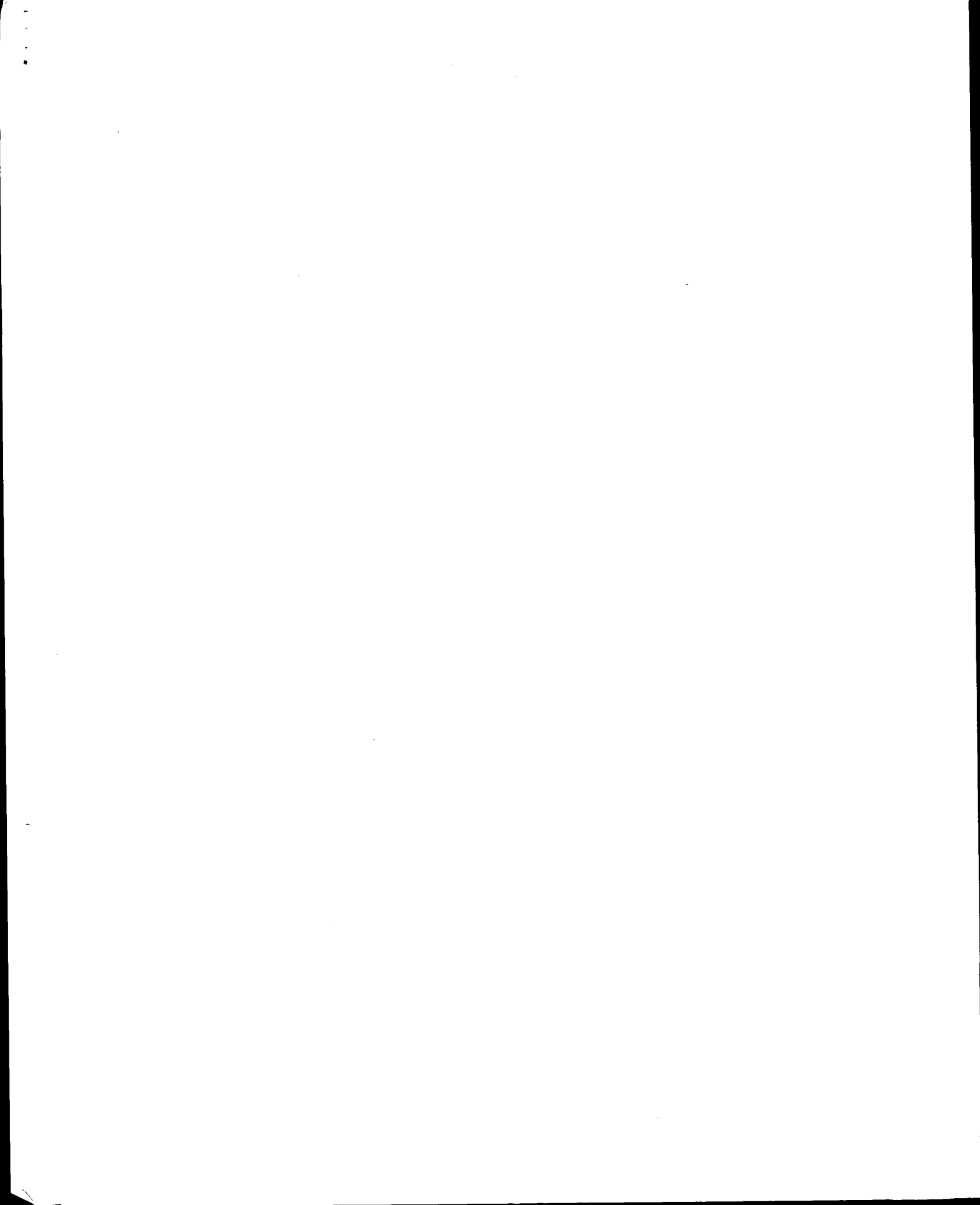
RESULT 15
AA217715
ID AA217715 standard; cDNA; 416 BP.
AC AA217715;
XX
XX 12-OCM-1999 (first entry)
DE Human gene expression product cDNA sequence SEQ ID NO:5188.
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;

```


Mon Aug 20 10:21:46 2001

us-09-284-320-56.rng

Page 17



© 2005 The Authors
Journal compilation © 2005 Blackwell Publishing Ltd

100

TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 523 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: BRSTUT01
; CLONE: 607227
US-08-628-413-1

Query Match 3.1%; Score 32.8; DB 2; Length 523;
Best Local Similarity 55.2%; Pred. No. 0.92;
Matches 64; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 459 gctccatgctgctgcttgaagaacgtgtctcagttcaccctcccaattctct 518
DB 272 gctccgctgagcctttttccagagcctgagctctgcttcttctggtgctcgt 213
QY 519 gagtagaacaatgaagttgacctctctcttcttgaactgcaagtgctacatg 574
DB 212 ctggcgaaccctggtctgacatcttcttcttggagctgctgacatcttctgag 157

RESULT 10
PCT-US96-12545-5/c
; Sequence 5, Application PC/US9612545
; GENERAL INFORMATION:
; APPLICANT: Wilson, Charles R.
; TITLE OF INVENTION: Expression Systems for Commercial
; TITLE OF INVENTION: Production of Cellulase and Xylanase in Bacillus subtilis
; TITLE OF INVENTION: and Bacillus licheniformis
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 40 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: PA
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/12545
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M5739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-832-2215
; TELEFAX: 610-941-6067
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5390 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Bacterial Plasmid DNA"
; IMMEDIATE SOURCE:
; CLONE: C4TET Plasmid Construct
PCT-US96-12545-5

Query Match 3.1%; Score 32.6; DB 5; Length 5390;

Best Local Similarity 55.9%; Pred. No. 3.6;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 280 attcgtaaccttgagaaatgagtccttctttagcttgacagtggtgcaattccatt 339
DB 1175 AATCCGTATCCTTACAGGATATTTGCGAGTTGCGCAATTCGCGATTTATCCGATT 1116
QY 340 cactctatcttcttgagaaactcctgtgttcttgcaagtgctccaccat 390
DB 1115 TATATTTATTTTTCGTCGCAATTTGACATTTTACATTTGAGATCATAGT 1065

RESULT 11
5304637-21/c
; Patent No. 5304637
; APPLICANT: DORSEY, LAMBERTUS C. J.; WAGEMAKER, GERARD; VOS,
; YONNE J.; VAN LEEUW, ROBERT W.; PERSON, MARIA L. N.
; TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF HUMAN
; INTERLEUKIN-3 AND MOTELINS THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/494,182
; FILING DATE: 13-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 249,184
; FILING DATE: 16-AUG-1988
; SEQ ID NO: 21:
; LENGTH: 7326
5304637-21

Query Match 3.1%; Score 32.6; DB 6; Length 7326;
Best Local Similarity 55.9%; Pred. No. 4.2;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 280 attcgtaaccttgagaaatgagtccttctttagcttgacagtggtgcaattccatt 339
DB 4924 AATCCGTATCCTTACAGGATATTTGCGAGTTGCGCAATTCGCGATTTATCCGATT 4865
QY 340 cactctatcttcttgagaaactcctgtgttcttgcaagtgctccaccat 390
DB 4864 TATATTTATTTTTCGTCGCAATTTGACATTTTACATTTTGGATCATAGT 4814

RESULT 12
US-08-418-085-3/c
; Sequence 3, Application US/08418085
; Patent No. 5869283
; GENERAL INFORMATION:
; APPLICANT: SLICKHUIS, HERMAN; SPULEN, GERARDUS CORNELIS
; TITLE OF INVENTION: PROCESS FOR OXIDATION OF STEROIDS AND
; TITLE OF INVENTION: GENETICALLY ENGINEERED CELLS USED THEREIN
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN & MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,085
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/054,185
; FILING DATE: 26-APR-1993

```

1 CLASSIFICATION: 435
2 PRIOR APPLICATION DATA:
3 APPLICATION NUMBER: US/08/002,608
4 FILING DATE: 11-JAN-1993
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: US/07/474,857
7 FILING DATE: 30-OCT-1990
8 PRIOR APPLICATION DATA:
9 APPLICATION NUMBER: US/07/474,798
10 FILING DATE: 16-JULY-1990
11 PRIOR APPLICATION DATA:
12 APPLICATION NUMBER: PCT/US89/00072
13 FILING DATE: 25-SEP-1989
14 PRIOR APPLICATION DATA:
15 APPLICATION NUMBER: NL/88/200904,6
16 FILING DATE: 06-MAY-88
17 APPLICATION DATA:
18 APPLICATION NUMBER: NL/88/202080,3
19 FILING DATE: 03-SEP-88
20 ATTORNEY/AGENT INFORMATION:
21 NAME: CHARLES A. MUSELIAN
22 REGISTRATION NUMBER: 19,683
23 REFERENCE/DOCKET NUMBER: 146,1169 CON-1
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: (212) 661-8000
26 TELEFAX: (212) 661-8002
27 INFORMATION FOR SEQ ID NO: 3:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 7336
30 TYPE: NUCLEIC ACID
31 STRANDEDNESS: DOUBLE
32 TOPOLOGY: UNKNOWN
33 FEATURE:
34 NAME/KEY:
35 LOCATION:
36 IDENTIFICATION METHOD:
37 OTHER INFORMATION: PLASMID pBHA-1
38
39 US-08-418-085-3
40
41 Query Match 3.1%; Score 32.6; DB 2; Length 7336;
42 Best Local Similarity 55.9%; Pred. No. 4.2;
43 Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0
44
45 Db 280 attcgagcccttgtagaagatgcagtcctttgctcttgacagtgtagaattccatt 339
46 4924 ATTCGATTCCTTACAGGATATTTCGAGTTGCTCAATTGCCATTGTAATCCGATT 4865
47
48 QY 340 cactcctattcttcgaggaactcctcgtgttttcagtggtcccaat 390
49 4864 TATTATTATTTTCGTCGATCATTTGACATTTTACATTGTGATCATCAAG 4814
50
51 RESULT 13
52 US-09-099-011A-3/c
53 ; Sequence 3, Application US/09099011A
54 ; Patent No. 6171836
55 ; GENERAL INFORMATION:
56 ; APPLICANT: SLUKHUIS, HERMAN; SELTEN,
57 ; APPLICANT: GERARDUS CORNELIS MARIA; SMAAL,
58 ; APPLICANT: ERIC BASTIAAN
59 ; TITLE OF INVENTION: PROCESS FOR OXIDATION OF
60 ; TITLE OF INVENTION: STEROIDS AND GENETICALLY ENGINEERED CELLS
61 ; NUMBER OF SEQUENCES: 79
62 ; CORRESPONDENCE ADDRESS:
63 ; ADDRESSEE: BIERMAN, MUSELIAN & LUCAS
64 ; STREET: 600 THIRD AVENUE
65 ; CITY: NEW YORK
66 ; STATE: NEW YORK
67 ; COUNTRY: USA
68 ; ZIP: 10016
69 ; COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,011A
FILING DATE: 17-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/418,085
FILING DATE: 06-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/054,185
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/002,608
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/474,857
FILING DATE: 30-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/474,798
FILING DATE: 16-JULY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/NL89/00072
FILING DATE: 25-SEPT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: NL88/200904.6
FILING DATE: 06-MAY-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: NL/88/202080.3
FILING DATE: 03-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 146,1169-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 7336 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: UNKNOWN
FEATURE:
OTHER INFORMATION: PLASMID PBHA-1
US-09-099-011A-3

Query Match      3.1%; Score 32.6; DB 4; Length 7336;
Best local Similarity 55.9%; Pred. No. 4.2;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0

QY 280 attcgacaccttgaggagaacgcagcttccttgcttgacagcagtggaacatcatt 339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4924 AATTCGATATCTTTTACAGATATTTTGCAGTTTCATTCATTCGCCATGATATATCCGATT 4865

QY 340 caactccattattctcgtggaactccctcgtgttgtagcagtggtcccaatt 390
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4864 TTTATTTATTTTTCGTCGATCATTTTGAACATTTCATATTTGCATCATAGT 4814

RESULT 14
US-08-470-369-21/c
; Sequence 21, Application US/08470369
; Patent No. 6238889
; GENERAL INFORMATION:
; APPLICANT: Dorsseers, Lambertus C. J.
; APPLICANT: Wagemaker, Gerard
; APPLICANT: Vos, Yvonne J.
; APPLICANT: Van Ieen, Robert W.
; TITLE OF INVENTION: MOLECULAR CLONING AND EXPRESSION OF HUMAN

```

```

? TITLE OF INVENTION: IL-3
? NUMBER OF SEQUENCES: 27
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Morrison & Foerster
? STREET: 755 Page Mill Road
? CITY: Palo Alto
? STATE: California
? COUNTRY: USA
? ZIP: 94304-1018
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/470,369
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/321,480
? FILING DATE:
? APPLICATION NUMBER: 07/854,297
? FILING DATE: 19-MAR-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Gracey, Nancy J.
? REGISTRATION NUMBER: 28,216
? REFERENCE/DOCKET NUMBER: 24615-20002.01
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-813-5600
? TELEFAX: 415-494-0792
? TELEX: 706141
? INFORMATION FOR SEQ ID NO: 21:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 7336 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: circular
? MOLECULE TYPE: DNA (plasmid)
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? IMMEDIATE SOURCE:
? CLONE: PBH41
? US-08-470-369-21

Query Match          3.1%; Score 32.6; DB 4; Length 7336;
Best Local Similarity 55.9%; Pred. No. 4.2;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 280 attcgtacccttggagaatgcagttcctttagcttgacagtggttgcgaattccatt 339
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 4924 AATCCGTATCCTTTACAGATATTTGCGAGTTTGCATTCATTTGCCATTGTATATCGCAT 4865
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 340 cactcctatttcttgaggaactcctgttcttgcagttgcagttgcctccagt 390
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 4864 TATATTTATTTTTCGTCGATCATTTGAACCTTTTACATTTGGATCATAGT 4814
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
US-08-460-343B-1/c
? Sequence 1, Application US/08460343B
? Patent No. 5741664
? GENERAL INFORMATION:
? APPLICANT: Marcus D. Ballinger and James A. Wells
? TITLE OF INVENTION: SUBSTITUTED VARIANTS CAPABLE OF CLEAVING
? TITLE OF INVENTION: SUBSTRATES CONTAINING BASIC RESIDUES
? NUMBER OF SEQUENCES: 74
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genentech, Inc.
? STREET: 1 DNA Way
? CITY: South San Francisco
? STATE: California
? COUNTRY: USA

```

```

? ZIP: 94080
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: WinPatIn (Genentech)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/460,343B
? FILING DATE: 01-Jun-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/398028
? FILING DATE: 03-Mar-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Koblenc, Jeffrey S.
? REGISTRATION NUMBER: 36,575
? REFERENCE/DOCKET NUMBER: P0936C1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650/225-8228
? TELEFAX: 650/952-9881
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 8119 base pairs
? TYPE: Nucleic Acid
? STRANDEDNESS: Single
? TOPOLOGY: Linear
? US-08-460-343B-1

```

```

Query Match          3.1%; Score 32.6; DB 1; Length 8119;
Best Local Similarity 55.9%; Pred. No. 4.5;
Matches 62; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 280 attcgtacccttggagaatgcagttcctttagcttgacagtggttgcgaattccatt 339
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5457 AATCCGTATCCTTTACAGATATTTGCGAGTTTGCATTCATTTGCCATTGTATATCGCAT 5398
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 340 cactcctatttcttgaggaactcctgttcttgcagttgcagttgcctccagt 390
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5397 TATATTTATTTTTCGTCGATCATTTTGAACCTTTTACATTTGGATCATAGT 5347
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: August 18, 2001, 21:53:05
Job time: 9227 sec

Mon Aug 20 10:21:38 2001

us-09-284-320-31.rni

Page 9

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 2001, 18:04:23 ; Search time 3117.45 Seconds
(without alignments)
3183.851 Million cell updates/sec

Title: US-09-284-320-31

Perfect score: 1050
Sequence: 1 atggctgtgttgcgtgc.....accagaagatcgaatgcat 1050

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1022815 segs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	853	81.2	1989	192	AK017482	AK017482 Mus muscu
2	654	62.3	781	154	BG530418	BG530418 602559166
3	634.2	60.4	733	32	AV721438	AV721438 AV721438
4	606.4	57.8	696	153	BG403007	BG403007 602418818
5	591	56.3	655	175	BG288867	BG288867 602383808
6	571.2	54.4	809	107	AU080111	AU080111 AU080111
7	567	54.0	947	141	BE987106	BE987106 601336663
8	565.4	53.8	779	154	BG529889	BG529889 602559899
9	560	53.3	624	122	AW598900	AW598900 EST370970
10	555.2	52.9	733	154	BG519469	BG519469 602577742
11	554	52.8	566	5	AA313317	AA313317 EST185213
12	552.2	52.6	1092	175	BG298183	BG298183 602396327
13	551.2	52.5	969	32	AV726680	AV726680 AV726680
14	550.4	52.4	962	174	BG173896	BG173896 602335955
15	545.2	51.9	586	8	AA528178	AA528178 np91h09.s
16	541	51.5	714	149	BF468315	BF468315 U1-M-CCO-
17	538	51.2	796	107	AU080227	AU080227 AU080227
18	534.4	50.9	629	32	AV702380	AV702380 AV702380
19	531.6	50.6	827	107	AU079816	AU079816 AU079816
20	529.4	50.4	787	107	AU080392	AU080392 AU080392
21	523.2	49.8	947	169	BF786430	BF786430 602112961
22	522.2	49.7	580	190	W37670	W37670 zc12803.r1
23	516	49.1	516	143	BF060652	BF060652 7159404.x
24	513.8	48.9	888	175	BG294880	BG294880 602392175
25	499.4	47.6	501	110	AM024459	AM024459 wu76106.x
26	499.4	47.6	852	154	BG501143	BG501143 602546382
27	496.2	47.3	957	153	BG426572	BG426572 602493487
28	495.8	47.2	994	175	BG298560	BG298560 602396945
29	493.8	47.0	762	166	BE312617	BE312617 601154595
30	493.4	47.0	943	174	BG172247	BG172247 602333391
31	489.6	46.6	486	7	AA397898	AA397898 zc72e08.s
32	485.2	46.2	490	102	AI817629	AI817629 wK3905.x
33	484.8	46.2	661	165	BE291891	BE291891 601084732
34	481.2	45.8	580	123	AM985491	AM985491 uE82c07.y
35	480	45.7	480	145	BF196461	BF196461 hr86e07.x
36	478	45.5	36	AM166105	AM166105 xf45c09.x	
37	473.8	45.1	688	112	BF101430	BF101430 601753236
38	473	45.0	582	159	NA1633	NA1633 y67f04.t1
39	471.4	44.9	939	137	BE569483	BE569483 601328304
40	471.2	44.9	861	168	BF697773	BF697773 602129945
41	468.6	44.6	868	144	BF100333	BF100333 601754021
42	468.2	44.6	968	144	BF119697	BF119697 601754021
43	460.6	43.9	751	155	BG570997	BG570997 602591585
44	454.6	43.3	843	151	BF655336	BF655336 602118683
45	454.4	43.3	897	145	BF178365	BF178365 601807810

ALIGNMENTS

RESULT 1
AK017482 1989 bp mRNA HTG 08-FEB-2001
LOCUS Mus musculus 8 days embryo cDNA, RIKEN full-length enriched
DEFINITION library, clone:5730403E06, full insert sequence.
ACCESSION AK017482
VERSION AK017482.1 GI:12856739
KEYWORDS CAP trappper.
SOURCE Mus musculus (strain:C57BL/6J) 8 days embryo cDNA to mRNA,
clone:11D:RIKEN full-length enriched mouse cDNA library
clone:5730403E06.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning

JOURNAL
REFERENCE
AUTHORS

Methods Enzymol. 303, 19-44 (1999)
2 (sites)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

TITLE

3 (sites)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T.,
Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T.,
Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, K., Ohara, E.,
Matsubara, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
Matsuda, S., Okazaki, Y., Muramatsu, M., Inoue, Y. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913

JOURNAL
MEDLINEREFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
REFERENCE
AUTHORS

FEATURES

source

1. 1989
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1902886"
/db_xref="MGI:1917745"
/clone="5730403E06"
/clone.lib="RIKEN full-length enriched mouse cDNA library"
/dev.stage="8 days embryo"
823. 1122
/note="putative"


```

Db      14 GGGGTCATGAGGAGGAGTGAACAACTGGCTCTACCCGAGCAGTGTGATTCGTACCT 73
QY      292 ttggagaatcagttccctttagcttgaacagttggaatccatccattt 351
Db      74 TTGGAGATGCAAGTTCCTTTTACGCTTGACAGTGTGCAAAATTCATTCACCTCTATT 133
QY      352 tctgagaatcctcctgttcttgaagttggtcccaagtgagaagatglatatgta 411
Db      134 TCTGAGGAACCTCCTGTTGTTTTCATGTTGGCTCCCACTGAGGAAGATGATATGTA 193
QY      412 gggagaagcaactcagtttgaagacccttgaacaccttgagccagctccgtaatgc 471
Db      194 GGGAGGCAACATCAGTGTGTTGAAGACCTTTACAGTACCTTGCGGCAAGCTCCGTATGTC 253
QY      472 ctgttcaagaagaactcgttctcagttcaactccctcaatctctcagtagagaacat 531
Db      254 CTGTTCTAAGAAACCTGTTCTCAGTTCACCTCCCTCAATTCCTGAGTAGAACAT 313
QY      532 gaattgacctcctcttcttcttgaactgcaagtgcaatgataatccaagcttctg 591
Db      314 GAAGTTGACCTCTCTTCTTCTTCTGAACTGCAAGTGTACATGATTTCAAGCTTGCTG 373
QY      592 tctcgtcaatagcatctagccaagatcatctcctgattatattcactgagctgca 651
Db      374 TCTGTCATTAACATCTAGCCAAAGATCATTCCTCTGATTTATTCACCTGAGCTGCA 433
QY      652 ggttgaatgaatggaagcgtttagggagaagctctgaacaatgagagatctct 711
Db      434 GGTTCGATGAATAATGGGAGACCTTATGGGAGACCTCTGAAACATTCAGAAATGCTTCT 493
QY      712 aagatcctgttgaagcgtctgcaaaagtttgagatgacatgacagctctatggtg 771
Db      494 AAGATCTGTTGAGCGCTGCAAAAGTTTGAGATGACATGATGATGATGATGATGATG 553
QY      772 aatgcaagtgatagatgtagtcaactgcaagtcatttgaacccctcattagaaga 831
Db      554 AATGCAAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 613
QY      832 aggaactatccttgagcaaaaagcaagaagccagcaagtcctcattacactgcatat 891
Db      614 AGGACTATCTTGGAGCAAAACAAAGCAAGCAAGCAAGTCCCTATTAACCTTGATAT 673
QY      892 aa 893
Db      674 AA 675

RESULT 3
AV721438 753 bp mRNA EST 16-OCT-2000
LOCUS AV721438 HTB Homo sapiens cDNA clone HTBALD05 5', mRNA sequence.
DEFINITION AV721438
ACCESSION AV721438
VERSION AV721438.1 GI:10818590
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 753)
AUTHORS Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, X., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu
S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z., and Han, Z.
Homo sapiens cDNA HTB clones
Unpublished (2000)
TITLE Contact: Zeguang Han
JOURNAL Chinese National Human Genome Center at Shanghai
COMMENT 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801922
Fax: 86-21-50801922
Email: hanzgechgc.sh.cn

```

```

FEATURES
Source This clone is available at CHGC in Shanghai.
Location/Qualifiers
1..753
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTBALD05"
/clone_lib="HTB"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 191 a 161 c 161 g 240 t
ORIGIN
Query Match 60.4%; Score 634.2; DB 32; Length 753;
Best Local Similarity 96.5%; Pred. No. 6e-183;
Matches 659; Conservative 0; Mismatches 23; Indels 1; Gaps 1;
QY 307 cctttagcttgaacagtgatgcaaatccattccattcttcttgaagaacacct 366
Db 25 CGTTTAGCTTGACAGTGTGCAAAATTCATTCACCTCTTATTCTTGAGGAACCTCT 84
QY 367 gtgtttgcaatggtcctccagtgagaagatglatatggttaggaagcaactca 426
Db 85 GTGTTTTCAGTGTGCTCCCACTGAGGAAGATGATGATGATGATGATGATGATGATG 144
QY 427 gtgttgaagaccttcaagtcacgttgcagcagctccgtaatgctgttccaagaag 486
Db 145 GTGTTTGAAGACCTTTAGTACCTTGGCCAGCTCCGTATATGCTGTTCAAGAAAC 204
QY 487 tctgttcagttcaactcccccctcaattctctgtagtagaacaatgaattgacctc 546
Db 205 TCTGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 264
QY 547 ttcttctgcaactgcaagtgatgacatgatttcaagctgtctcgtcgtcataagac 606
Db 265 TTTCTTCTGAACTGCAAGTGTACATGATTTTCAAGCTTGTCTGTCATTAACAT 324
QY 607 ctgagcaagatcatctcctgattatattcactggaagctggaagttgagatgaat 666
Db 325 CTAGCCAAAGATCATTCCTCTGATTTATTCACCTGAGCTGAGCTTGTGATGAAT 384
QY 667 gggagaagcgtttagggagaagactctgaacaattcagaagatgcttcaagctgttga 726
Db 385 GTTAACCGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 444
QY 727 gctctgcaaaagtttgagatgacatgacatgcttattatggttaggaatgcaatg 786
Db 445 GCTCTGCAAAAGTTTGAGATGATGATGATGATGATGATGATGATGATGATGATG 504
QY 787 ttatgcaatgcaagtcatttgaacacccctcattaggaagaagacacacacctgag 846
Db 505 TTATGACCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 564
QY 847 ggaagaagcaaggaagcaagcaagctccctataacacctgcatataagataattgga 906
Db 565 GCAAAATGAGGAATTAACCAATCAAGTCCCTATTAATTCATTAATTAATTTTGA 624
QY 907 taatcgtgttcttcaacatgta-cttgataatgacgctgagctgagcttgaat 965
Db 625 TAATCCGCGCTTCAACATGACCATGATGATGATGATGATGATGATGATGATGATG 684
QY 966 tatcaacctcaacatattgga 988
Db 685 TATCACTCTTCAATATTGAA 707

RESULT 4
BG403007 996 bp mRNA EST 12-MAR-2001
LOCUS BG403007
DEFINITION 602418818F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4525781 5',

```

mRNA sequence.
 ACCESSION BG403007.1 GI:13296455
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/
 1 (bases 1 to 996)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaps-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLAM10432 row: d column: 06
 High quality sequence stop: 702.
 Location/Qualifiers
 1. 996
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4525781"
 /clone_lib="NIH_MGC_93"
 /tissue_type="transitional cell papilloma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Bladder; Vector: PCMV-SPORE6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
 BASE COUNT 279 a 186 c 235 g 296 t
 ORIGIN
 Query Match 57.8%; Score 606.4; DB 153; Length 996;
 Best Local Similarity 96.3%; Pred. No. 2.2e-174;
 Matches 676; Conservative 0; Mismatches 16; Indels 10; Gaps 5;
 Oy 346 tttattctgaagaactcctcgtgttctgacgtgtgctccagtaggaagaagtgat 405
 1 tttattctgaagaactcctcgtgttctgacgtgtgctccagtaggaagaagtgat 60
 Oy 406 atgtaggaagaagcaactcagtggttgaagaagccttcaagccttgcgcagctcgt 455
 61 atgtaggaagaagcaactcagtggttgaagaagccttcaagccttgcgcagctcgt 120
 Oy 466 aatgcctgttcaagaagaactcgtctcagttcaactcctccctcaattctcgtag 525
 121 aatgcctgttcaagaagaactcgtctcagttcaactcctccctcaattctcgtag 180
 Oy 526 aacaatgaagttagcctccttcttcttgaactgcgaagtgctcaatgatttcaagc 585
 181 aacaatgaagttagcctccttcttcttgaactgcgaagtgctcaatgatttcaagc 240
 Oy 586 ttgctgtcgtcaataagcatcagcgaagaatcattcctcgtattatattcactgag 645
 241 ttgctgtcgtcaataagcatcagcgaagaatcattcctcgtattatattcactgag 300
 Oy 646 ctggcaggtttagatgaattgggaagcgttatgggaagccttgaacaattcagaat 705
 301 ctggcaggtttagatgaattgggaagcgttatgggaagccttgaacaattcagaat 360
 Oy 706 gctctaaagatcctgttgaagcctcgaagaatttgaagaatgacatgacatgattat 765
 361 gctctaaagatcctgttgaagcctcgaagaatttgaagaatgacatgacatgattat 420
 Oy 766 ggttggaatgagtgtagtgtagtcaatcgtcgaagtcatttgacacccctccatag 825

Db 421 ggttggaatgagtgtagtgtagtcaatcgtcgaagtcatttgacacccctccatag 480
 Oy 826 aagacaagacatccttgaagcaagaagcgaagaaccagaaatccatata--cc 883
 Db 481 aagacaagacatccttgaagcaagaagcgaagaaccagaaatccatata--cc 540
 Oy 884 ttgcataagta-taatttgaatcctgag--tttcaacatgtaacttgaata 940
 Db 541 ttgcataagta-taatttgaatcctgag--tttcaacatgtaacttgaata 600
 Oy 941 tgaatgccttgagcctgctgctgattatcacccttcaacatatttgaagaatgactg 1000
 Db 601 tgaatgccttgagcctgctgctgattatcacccttcaacatatttgaagaatgactg 657
 Oy 1001 gataatgag--catcattatagatgataacacagaagat 1040
 Db 658 gataatgag--catcattatagatgataacacagaagat 699
 RESULT 5
 BG288867 665 bp mRNA EST 21-FEB-2001
 LOCUS 602383808F1 NIH_MGC_93 Homo sapiens CDNA clone IMAGE:4513009 5',
 DEFINITION mRNA sequence.
 ACCESSION BG288867 GI:13044136
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/
 1 (bases 1 to 665)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaps-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLAM10398 row: p column: 02
 High quality sequence stop: 665.
 Location/Qualifiers
 1. 665
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4513009"
 /clone_lib="NIH_MGC_93"
 /tissue_type="transitional cell papilloma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Bladder; Vector: PCMV-SPORE6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
 BASE COUNT 185 a 128 c 145 g 207 t
 ORIGIN
 Query Match 56.3%; Score 591; DB 175; Length 665;
 Best Local Similarity 100.0%; Pred. No. 9.8e-170;
 Matches 591; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 460 ctccgaatgcctgttcaagaagaactcgtctcagttcactcctccatattctct 519
 1 ctccgaatgcctgttcaagaagaactcgtctcagttcactcctccatattctct 60
 Oy 520 agtagaacaatgaagttagcctgtcttcttctgaactgcaagtgctacatgataatt 579


```

DEFINITION 601439663F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3924687 5',
            mRNA sequence.
ACCESSION  BE897106
VERSION    BE897106.1 GI:10362234
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
          Email: cgsaps-r@mail.nih.gov
          Tissue Procurement: ATCC/DCMP
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Plate: L1AM9762 row: n column: 16
          High quality sequence stop: 656.
FEATURES
  source
    1..947
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:3924687"
    /clone_1lb="NIH_MGC_72"
    /tissue_type="melanotic melanoma"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: skin; Vector: pCMV-SPOK6; Site:1: NotI;
    Site:2: SalI; Cloned unidirectionally. Primer: Oligo dF.
    Average insert size 2 Kb. Library constructed by Life
    Technologies."
BASE COUNT  274 a 155 c 210 g 307 t 1 others
ORIGIN
Query Match 54.0%; Score 567; DB 141; Length 947;
Best Local Similarity 100.0%; Pred. No. 2.6e-162;
Matches 567; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 484 aacctgctctagctcactcccccctcaattctcgcagtagaacaatgaaatgacctg 543
DB 1 AACCTGCTCTCAGTTCACCTCCCTCAATCTCTGAGTAGAACAATGAAAGTTGACCTG 60
QY 544 ctcttccttcgaaatgcaagtgctgaatgatttcaagcttgctgctgctcaatga 603
DB 61 CTCTTCCTTCGAACTGCAAGTGTCTACATGATATTTTAAAGCTTGCTGCTGATTAAG 120
QY 604 caltagcaagaatcaltctcctgattatatactcagtcgagctggcaggttgatgaa 663
DB 121 CATCTAGCCAGAGATCATTTCTCGATTATATTCATGAGCTGCCAGCTTGGATGAA 180
QY 664 attggaagagttatgggaagagctctgaaacattcagaagtgctcttaagatcctggt 723
DB 161 ATTGGAGAGCGTTATGGGAGAGACTCTGAAACAATTCAAGATGCTTCTAAGATCTGTT 240
QY 724 gacgctcgtgaaagtctgacagatgacatgacgtccttaattggtggaaatgacgtgta 783
DB 241 GACGCTCTGCAAAAGTTTGCAGATGACATGTACACTCTTATGTGTGGGAATGACATGCTA 300
QY 784 gaattgaatcagtaagtaattgtaacacatccctcatatgaagaagaagatcctt 843
DB 301 GAGTTTATGCTACTGATCACTTTGTAACCTCCCTCATTTAGAAAGCAAGACATTCCTT 360
QY 844 gaggaacaacaagaagaaacccagcagatccctataaacttgcatataatgaatttc 903
DB 361 GAGGCAAAACCAAGCAGAGAACCCAGAAAGTCCCTATTAACCTTGAATTAATTAATT 420
QY 904 gaatatccgttggtttcaacatgctacttgataaataatgacgccttgacctgctg 963

```

```

DB 421 GAATATTCCTGCTGTTTCAACATGCTACTTTGGATATATGATGCTTGGCTGCTG 480
QY 964 attatccctcttaacaatatttgaacaatgacatcctgagatgataatgacatcattag 1023
DB 481 ATTATCACCCTCTTCAATATTTGGAAACATGATCTCGATGATGATGATGATTTATAG 540
QY 1024 attacaacaagaagattcgaatgatt 1050
DB 541 ATGACAAACCAAGATTCATGATGAT 567

RESULT 8
BG529889 779 bp mRNA EST 03-APR-2001
LOCUS 60258989F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:469630 5',
DEFINITION mRNA sequence.
ACCESSION BG529889
VERSION BG529889.1 GI:13521426
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
          Email: cgsaps-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: CLONETECH Laboratories, Inc.
          DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Plate: L1CM1524 row: b column: 23
          High quality sequence stop: 739.
FEATURES
  source
    1..779
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:469630"
    /clone_1lb="NIH_MGC_61"
    /tissue_type="embryonal carcinoma"
    /lab_host="DH10B (T1 phage-resistant)"
    /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site:1:
    SfiI (ggccgctcgcc); Site:2: SfiI (ggccatagcc);
    double-stranded cDNA was prepared from cell line RNA. 5'
    and 3' adaptors were used in cloning as follows: 5'
    adaptor sequence: 5'-CACGGCCATTAAGGCC-3' and 3' adaptor
    sequence: 5'-ATTCTAGAGCCGACGCCGCCGACATG-3' (30)BN-3'
    (where B = A, C, or G and N = A, C, G, or T). Average
    insert size 1.75 Kb (range 0.9-4.0 Kb). 15/15 colonies
    contained inserts by PCR. This library was enriched for
    full-length clones and was constructed by Clontech
    Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
    Library."
BASE COUNT  157 a 198 c 208 g 216 t
ORIGIN
Query Match 53.8%; Score 565.4; DB 154; Length 779;
Best Local Similarity 97.5%; Pred. No. 7.3e-162;
Matches 627; Conservative 0; Mismatches 11; Indels 5; Gaps 5;
QY 1 atggcgtgttgctgctcctgctgctgctgctgctgctgctgctgctgctgctgctgct 60
DB 137 ATGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 195
QY 61 atataaataccacagagctgctgctgcttccgaaatggaaatggcctataccagagag 120

```

Db 196 ATATTAATACACGAGGCTGCTGTTTCCGAATGGAATTGGCTTACACGAGAG 255

QY 121 cggatccagacgctgctgcatctgcccagtgagcctcctcgtgaaagacccctcttg 180

Db 256 CGATCCACGAGAGCTGGCTGCTGCTCATGGGCTTCTCTGAAAGAACACCTTCTTGG 315

QY 181 ccagagctcgacg-ctggtaacctgttcatctgctcccgagctacacgcatctgagatgct 239

Db 316 CAGAGACTCGACGCTGGGTAACTGTTTATGCTGCTGCGGCTACCGCTATGCTGATGCT 375

QY 240 gaaggagtgacaaacacgcttaccgcccaaggagtgatcttctgaccccttgagaa 299

Db 376 GAAGGAGTGAACAACATGCTCTACCCACGACGATGTCATTTCCTACCTTGGAGAA 435

QY 300 tgcagctcttcttctgctgacagtgctgacaaatcttacccttcttcttgagaa 359

Db 436 TGCAGTCTCTTATGCTGCTGACGCTTTCGAATTCATTCATCTTATTTTCTGAGGA 485

QY 360 aactcctgttcttctgacgttgcctccagtgagaa-agagtgtatagtgagaaag 418

Db 496 AACTCTGTTGTTTGGCACTTGGCTCCAGTGAAGAACAGAGTGTATAGTGAAGG 555

QY 419 caaacctgcttctgacagcttctgacacccctgagcagctccgtaatgccc-tgtt 477

Db 556 CAACCTCATGTTTGAAGACCTTTCAGTACCTTGGCCGACCTGCTAATGCTGCTT 615

QY 478 caagaaacactgttctgacgttctcctccctcaactctctgagtagaacaatgag 537

Db 616 CAAGAAACCTGTTTCTGCTGCTGCTCCCTCAATCTCTGAGTGAAGAACATGAAGT 675

QY 538 gacctgtcttcttcttctgacagctgacagtgacatgacatcttcaagcttgc-gtctg 596

Db 676 TGACCTGTTTCTTCTTCTGACAGTGAAGAGCTACATGATATCCAGAGCTGCTGCTG 735

QY 597 tcatagacatcagcaggaatcatcttctctgattatca 639

Db 736 TCATTAAGCATCTAGCCAAAGATCATCTTCTGATTATTATCA 778

RESULT 9
AM958900 624 bp mRNA EST 01-JUN-2000
LOCUS
DEFINITION EST170370 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION AM958900
VERSION AM958900.1 GI:8148584
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 624)
AUTHORS Hegde, P., Qi, R., Aherathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
COMMENT Unpublished (2000)
CONTACT: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 127
Seq primer: Reverse.

FEATURES
source Location/Qualifiers
1..624
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGE"
/note="Vector: pBluescriptSkm"
BASE COUNT 162 a 135 c 134 g 193 t
ORIGIN

Query Match 53.3%; Score 560; DB 122; Length 624;
Best Local Similarity 97.1%; Pred. No. 3e-160;
Matches 602; Conservative 0; Mismatches 15; Indels 3; Gaps 3;

QY 278 tcatctgacaccccttggagaaatgacagctcttcttctgacagtgctgcaatctca 337

Db 1 TCATTCGTACCTTTTGGAGAAATGCAAGTTCCTTTAGCTTTCAGACATGTTGCAATTC 60

QY 338 ttaactcctatcttctgagaaacccctgttcttctgacagtgctccagtgagaa 397

Db 61 TTCACCTCTTATTTTTCGAGAAACCTCTGTTGTTGCAATGGCTCCAGTAGGAA 120

QY 398 gaatgtatattgtaggagaaagcaactagtggttgaagaccttcaagctctgagc 457

Db 121 GAGTGTATATGTTAGGAGAGCAAACTCAGTGTGGAACCTTTCAGTACCTTGGCC 180

QY 458 agctccgtaatgcctgttcaagaaacactgttctcaagttcaactcccccatactc 517

Db 181 AGCTCGTATGCTGCTGTTTCAGAAACCTGTTTCAGTTCATCTCCCTCAATTC 240

QY 518 tgaatgagaaatgaaagtgacagcttcttcttcttctgacagtgagtgacatgaa 577

Db 241 TGAATGGAACAATGAAGTTTACCTCTCTTCTTCTGAACTGCAAGTGTACATATA 300

QY 578 ttcaagctgtctgtctgcataaagcatcagcagagacatctctctgattatatt 637

Db 301 TTTCAAGCTTGTGCTGCTGCTGCTATACATCTAGCAAGATCTCTCTGATATATT 360

QY 638 cactgagactgacaggttctgagaaatgagagcgttatggagaaactctgaacat 697

Db 361 CACTGAGCTGCGCAGGTTTGGATGAAATTTGGAGCGTTATGGGGAACACTGTGAACA 420

QY 698 tcagagatcctctcaagatcctgttgaagctctgcaaaagttgcaatgacatgaa 757

Db 421 TCAGAGATGCTCTTATAGTCTTGTGAGCGCTGCAAAAGTTTGAGATGACATGAC 480

QY 758 gtctttagtgtaggaatgacagtgtagtagtagtcaactgtcaagttcaactctcc 817

Db 481 GTCTTATGTTGGAGATGAGTGTAGTACTGTCAAGTCAATTTTACACCTCC 540

QY 818 tcatagagaaacagacatctcttggagcaaaacagagaaacccagaaactctcc 877

Db 541 TC-TTAGGAAGCAAGACATTATCTTGAAGCAAAAC-AGCAAGAAACCA-CCAGTCCG 597

QY 878 ataactgtacataagtat 897

Db 598 ATACTTGTCTATTAAGATTAAT 617

RESULT 10
BG519469 733 bp mRNA EST 07-APR-2001
LOCUS
DEFINITION BG519469 NC1_CCAP_Mam6 Mus musculus cDNA clone IMAGE:3497818 5',
RNA sequence.
ACCESSION BG519469
VERSION BG519469.1 GI:13515113
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 733)
AUTHORS NIH-WGC <http://wgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-f@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

OY 114 agagagagcagatcccaacatgctgctcattgcatggtctctctgtgaaagaagact 173
 |||
 DB 61 AGAGAGCGGATCCAGACGAGGCTCATGTCATGAGGCTCTCTGTGAAAGAACCT 120
 |||
 OY 174 tctctggcagagactcgaatgggtaacctgttcaatcgtcccgagctacgctcagc 233
 |||
 DB 121 TTCCTGGCAGGACTCCGAGTGGGTAACTGTTTATGTCGTCGAGGCTACGCTCATG 180
 |||
 OY 234 gatggtgaagggatggaacaaatgcttaccocccagcagatgctatctgtaacctt 293
 |||
 DB 181 GATGCTGAAGGAGGAGAACAACTGCTTACCCCGAGGAGTGTATTTGCTACCTTT 240
 |||
 OY 294 ggaagatcagctcttcttctgacagtgctgacaaatccatccatctatcttctc 353
 |||
 DB 241 GGAGATGCAATGCTTTTGTAGTCTTACAGTGTGCAATTCATTCATCTTATTTTC 300
 |||
 OY 354 tgaagaaactccttcttctgacagtgctcccaagtggaagaaagtgatagtagg 413
 |||
 DB 301 TGAGGAATCTCTGTTTGTGAGTGGCTCCAGTGAAGAAAGTGTATATGATAGG 360
 |||
 OY 414 gaaagcaactcagtgcttgaagaccttcagcttcagcttcgcaagctcgaatcgc 473
 |||
 DB 361 GAAGGCAAACTCAGTGTGTAAGACCTTTGAGTCACTTGCAGCA-NTCCGTAATGCGCT 419
 |||
 OY 474 gttcaagaaactcgttctcagctacccctcaatctctcagataggaacaatga 533
 |||
 DB 420 GTTTCAGAAACTCTGTCTCAGTTCACCTCCCTCAATTCCTGAGTGAACAAATGA 479
 |||
 OY 534 agttgacctctcttcttctcagactcgaagtgctcagatataatcaactcgtctc 593
 |||
 DB 480 AGTTGACCTGCTCTTCTTCTGACTGCAAGTGTCTCATGATATTCACCTTCTGCTG 539
 |||
 OY 594 tctcaataagcactagcgaagatca 620
 |||
 DB 540 TCCTCATTAAGCATCTACCCAAAGATCA 566
 |||

RESULT 12
 BG298183 1092 bp mRNA EST 21-FEB-2001
 LOCUS 602396327F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4507782 5'
 DEFINITION mRNA sequence.
 ACCESSION BG298183
 VERSION BG298183.1 GI:13062399
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1092)
 NIH-MGC http://mgi.mcl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM10385 row: f column: 07
 High quality sequence stop: 710.
 Location/Qualifiers
 1. 1092
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:4507782"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: PCMV-SPOB6; Site: 1; Note:
 Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH-MGC Library."
 BASE COUNT 313 a 262 c 261 g 256 t
 ORIGIN

Query Match 52.6%; Score 552.2; DB 175; Length 1092;
 Best Local Similarity 89.5%; Pred. No. 9.3e-158;
 Matches 606; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

OY 374 tgaagtggctccacagaggaagaaagtgatagtgaggaagcaactcagtgcttg 433
 |||
 DB 1 TGCAGTTGGCTCCAGCAGAGAGAGAGTGTATATGTTGGGGAACCAACTGGGTGG 60
 |||
 OY 434 aagaccttcagctaccttcgcaagctcgaatcgcgtcttcaagaagaactcgttc 493
 |||
 DB 61 AAGACCTTCAGTCAATTCGCGCAGACCTCCGTAACCCGCTGTTCAAGAAATCCCTTC 120
 |||
 OY 494 tcaatcactccctcctcaatctctcagataggaacaatgaagtgaactgtcttctt 553
 |||
 DB 121 TCAACTCTCTCCCTCAATTCCTGAGTGAAGTATGAGTGAACCTGCTCTTCTCT 180
 |||
 OY 554 ctgaactcgaagtgctcagatataatcaagcttgcgtctcgtcatalaagcatcagca 613
 |||
 DB 181 CCGAATCGCAAGTCTCATGATATTTCCAGTTGTTGTCGTCATGAGCATCGCA 240
 |||
 OY 614 aggaatctccctcagatataatcaagctcgaagctcgaagcttggatgaattggagaac 673
 |||
 DB 241 AGACCATTCACCCGACTTGTATTCATGAGAGTGGCAGGTTGGATGAACCTTGGAAC 300
 |||
 OY 674 gtatagggaagactcgaacaattcaagatgcttctcagatcccttctgagctctgc 733
 |||
 DB 301 GTTATGGGAGAGACTGAGACAGTTCAGAGATCCTCTAAGATCCTGTGATGCTCTCC 360
 |||
 OY 734 aaaaatttcagatgacatgctcagcttcttctgaggaatcagtgatagtgatga 793
 |||
 DB 361 AAAAGTTTCAGATGACATGTACATGCTCTATGAGTGGAGACGAGGTGATGATGAGTGA 420
 |||
 OY 794 ctgtaagctatctgacacccctccatcaggaagaagaactatcccttgaggcaaac 853
 |||
 DB 421 CTGTCAAAATCATTTGACACATCTCCCTGTGAGGAGTCAAGCAATCCTTGAGCAAAAC 480
 |||
 OY 854 aagcgaagaccagaagcctcctaactcgtcagatataatgaattgaattctcg 913
 |||
 DB 481 AAG---AGACACCCCAAGTCCCTATTAACCTTGATTAATGATTTGAGTATTCAG 537
 |||
 OY 914 tggtttcaacatgacttggatataatcagcttggccttggccttgcttgatataacct 973
 |||
 DB 538 TGGTTTCAACTGTGACTGTGATTAATGATGAGCTTGAGCTTGCTGTGATTAACCT 597
 |||
 OY 974 cttaaatatttgaacatgagctcgtgataatagcattatagatagaacaac 1033
 |||
 DB 598 CTTCAACAATTTGGAATATGATCCTGGATATGATGACATCATATTAAGATGACAAATC 657
 |||
 OY 1034 agaagattcgaatgat 1050
 |||
 DB 658 AGAAGATCGAATGAT 674
 |||

RESULT 13
 AV726680
 LOCUS AV726680 969 bp mRNA EST 17-OCT-2000
 DEFINITION AV726680 HMC Homo sapiens cDNA clone HMCAB11 5', mRNA sequence.
 ACCESSION AV726680
 VERSION AV726680.1 GI:10836101
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

Db 308 ATGCAAGTTCCTTTAGCTAGACAGCGTGGCAATTCATTCTCTTATTTTCGGAG 367
Qy 358 gaacacctgtgttttgcagttgctccagtgaggaagatglatatgtaggaag 417
Db 368 GAAACCTGATGTTTGAAGACCTTTCAGTCAATTTGGGAGCTCCGTACCGCTTTT 427
Qy 418 gcaaacctgagtttgaagaccttcaagtcaccttgcgcagctccglaacgcctgtt 477
Db 428 GCAAACTCGGTGTTTGAAGACCTTTCAGTCAATTTGGGAGCTCCGTACCGCTTTT 487
Qy 478 caagaaacctgtgttccagtttcaatccctcaatcttctgaagaaacatgaagt 537
Db 488 CAAGAAACCTCCCTTCAACTCTCTCCCTCAATCTCTGAGTAGAATATAGAGCTT 547
Qy 538 gaacctgtcttcttcttgaagtcagtcagtcagtcagtcagtcagtcagtcagtc 597
Db 548 GACCTGCTCTTCTCTCCGAACTGCAAGTGTACATATTTCCAGTTTCTCTCGCT 607
Qy 598 ctaagacatcagccaagatcattctcctgattatlatcact-gaagctgcaagtt 656
Db 608 CATAGCACTAGGCC-AGGACCATTCACCGACTTGTATTCATTGGAGCTGCGAGCTT 666
Qy 657 ggaatgaatttggaagcgttat-999gaagacatcgaacatcgaagatgcttcaag 714
Db 667 GGAATGACTGTGGAAGCGTATGCGGGGAGACATCTGAACCACTTCAGATGCTCTTAAG 726
Qy 715 atcctgttgcagctctgcgaagattgcagatgacatgacatgacatgacatgacat 774
Db 727 ATCTGTGATGCTCTCCCAAGATTTGAGTACCTGTACAGCTGCTCTTATAG-GGGAGC 785
Qy 775 gcagtgagtagtagttagtcaagtcagtcagtcagtcagtcagtcagtcagtcagtc 834
Db 786 GCGAGGCTCAGAG-TAGTGAAGTGGCAAAATTTCTGGGAACCATTCCTGTGAGAAATGAG 844
Qy 835 actactccttgaggaagaaagcgaag 860
Db 845 ACATTTCTTGAAGAAAGAACCA 870

RESULT 15
AA528178 586 bp mRNA EST 05-AUG-1997
LOCUS nb91h09.s1 NCL_CGAP_Brl.1 Homo sapiens cDNA clone IMAGE:965921 3'
DEFINITION mRNA sequence.
ACCESSION AA528178
VERSION AA528178.1 GI:2270247
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 586)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 2663 Std Error: 0.00
Seq primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 390.
Location/Qualifiers
1. .586
FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="IMAGE:965921"
/clone_lib="NCL_CGAP_Brl.1"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is not normalized. (The normalized
version of this library is NCI CGAP Br2.) Library was
constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 51.9%; Score 545.2; DB 8; Length 586;
Best Local Similarity 96.8%; Pred. No. 1e-155;
Matches 567; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

Qy 46 ggaagcaggtttatataataaacaccaggtctgtgttccgaatggaattg 105
Db 1 GGGAGCGAGTTTATATATTAAATCACCAGGCTGTGTTTCCGAATGGAATTTGG 60
Qy 106 cctatacagaagagcgatcccaagcgtgctgcatgtlccatgtggtctctgtgaa 165
Db 61 CTTATACAGAGAGCGAGATCCAGACGTGCTGCAATGTGCGATGGGCTTCTGTGAA 120
Qy 166 gaagacgttttggcgaagactcgaatggtgaaccttcaatgctcgtgggtacc 225
Db 121 GAAGACCTTCTTGGCCAGAGCTCGCACTGGGTAACCTGTTTCATGCTCTCGGGTACC 180
Qy 226 gtcactgtagtctgtagaaggaatgaaacaactgctctacccccagagagtgatctcg 285
Db 181 GTCATGATGATGTGAAGGAGTGAACAATGAGCTTACCCCGAGAGTGCATTTGG 240
Qy 286 tacccttggagaatcagttcttctttagtgcagtgatggaatcattccatcc 345
Db 241 TACCTTGGAGATGCAAGTCTCTTGTGATGAGTGTGCAAAATTCATTCTCTC 300
Qy 346 ttatttctgagaactctctgttcttgcagttgctcccaagtgaagaaagtat 405
Db 301 TGATTTCTGAGAACTCTGCTGTGTTGCAATGGCTCCAGTGAAGAAAGTGTAT 360
Qy 406 atgtaggaaggaactcaatggttgaagaccttcaatgtaaccttgccagctcgt 465
Db 361 ATGTAGGGAAGCAACTCACTGTTGAAGACCTTTCAGTACCTTGGCCAGCTCGT 420
Qy 466 aatgcctgttcaagaactctgttctcagttcactcccccctaattctcgaatag 525
Db 421 AATCCCTGTTTCAAGAAACTCTGTGCTAGTTCAATCCCTTCATTTCTGAGATAG 480
Qy 526 aacaaagaagttagcctgtcttcttctcgaatgcaagtgcaatgatattcaagc 585
Db 481 AACAAATGAGTGAAGCTCTCTTGTGTGAAGCCGAATGGAATGAGATTTCAGC 540
Qy 586 ttgctgtctgcatagaatcagcca-ggatacttctcgtat 630
Db 541 TTGCCGCTGCTCCAAAGCATCTAGCAAGGAGTATTTCCCTGAT 586

Search completed: August 18, 2001, 20:15:30
Job time: 7867 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:15:33 ; Search time 22.94 Seconds
(without alignments) 924.952 Million cell updates/sec

Title: US-09-284-320-6
Perfect score: 1763
Sequence: 1 MAVFVLLALVAGVIGNEFS.....MDPGYSLIYRMTNQRKIMD 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 segs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1763	100.0	350	19 AAW64539	Human epidermoid c
2	1763	100.0	350	22 AAB90553	Human secreted pro
3	1757	99.7	350	22 AAB88347	Human membrane or
4	1751	99.3	350	22 AAB90589	Human secreted pro
5	1486	84.3	335	20 AAV59720	Secreted protein 3
6	441	25.0	93	20 AAV13118	Human secreted pro
7	121.5	6.9	364	21 AAG11429	Arabidopsis thalia
8	121.5	6.9	392	21 AAG11428	Arabidopsis thalia
9	103.5	5.9	428	19 AAW98378	H. pylori GHPD 137
10	101.5	5.8	318	21 AAG11430	Arabidopsis thalia
11	101	5.7	412	19 AAV10994	H. pylori ORF 09ce

12	100.5	5.7	1103	14 AAR39631	Neurofibromatosis
13	100.5	5.7	2485	15 AAR59921	RAS associated GAP
14	100.5	5.7	2485	15 AAR59922	RAS associated GAP
15	100.5	5.7	2818	13 AAR22268	NTI gene product.
16	100.5	5.7	2818	13 AAR22268	Human neurofibromi
17	97.5	5.5	982	21 AAB10105	Feline foamy virus
18	97	5.5	428	21 AAV28433	Corn leuc subunit
19	96.5	5.5	1886	21 AAW54241	Rattus norvegicus
20	96	5.4	853	21 AAG47981	Arabidopsis thalia
21	96	5.4	916	21 AAG47980	Arabidopsis thalia
22	96	5.4	922	21 AAG47979	Arabidopsis thalia
23	93	5.3	2627	19 AAG61347	Human telomerase R
24	92	5.2	451	21 AAG40227	Arabidopsis thalia
25	92	5.2	469	21 AAG40226	Arabidopsis thalia
26	92	5.2	517	21 AAG40225	Arabidopsis thalia
27	92	5.2	595	19 AAW98283	H. pylori GHPD 125
28	92	5.2	908	20 AAV33451	Chimeric Tag DNA P
29	91.5	5.2	454	21 AAG53794	Arabidopsis thalia
30	91.5	5.2	511	21 AAG53793	Arabidopsis thalia
31	91.5	5.2	545	21 AAG53792	Arabidopsis thalia
32	91.5	5.2	789	19 AAW46868	Bacillus thuringie
33	91.5	5.2	789	20 AAV24971	Bacillus thuringie
34	91	5.2	448	21 AAV84831	A. Streptococcus pn
35	91	5.2	457	21 AAV84830	A. Streptococcus pn
36	91	5.2	3066	18 AAW36178	Murine Ataxia-tela
37	90.5	5.1	349	19 AAW44368	Aspergillus nidula
38	90.5	5.1	438	21 AAG19277	Arabidopsis thalia
39	90.5	5.1	454	21 AAG19276	Arabidopsis thalia
40	90.5	5.1	454	21 AAG43053	Arabidopsis thalia
41	90.5	5.1	546	21 AAG43052	Arabidopsis thalia
42	90.5	5.1	561	21 AAG43051	Arabidopsis thalia
43	90.5	5.1	611	21 AAV66643	Membrane-bound pro
44	90.5	5.1	673	21 AAB38323	Human secreted pro
45	90.5	5.1	673	21 AAB07428	Amino acid sequenc

ALIGNMENTS

RESULT 1
ID: AAW64539 standard. Protein: 350 AA.
XX
AC AAW64539:
XX
DT 21-OCT-1998 (first entry)
XX
DE Human epidermoid carcinoma cell line KB clone HP10034 protein.
XX
KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW haematopoiesis; activin; inhibitor; chemotactic; chemokine; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.
XX
OS Homo sapiens.
XX
PN WO9821328-A2.
XX
PD 22-MAY-1998.
XX
PF 07-NOV-1997; 97WO-JP04056.
XX
PR 13-NOV-1996; 96JP-0301429.
XX
PA (PROT-) PROTECNE INC.
XX (SAGA) SAGANT CHEM RES CENTRE.
XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX WPI: 1998-297932/26.
XX N-PSDB: AAV49560, AAV49561.
XX Human protein having transmembrane domain - useful for, e.g.

PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA,

KW Human; secretory protein; membrane protein; vaccine; gene therapy;

KM	rheumatoid arthritis; diabetes.
XX	
XX	
OS	Homo sapiens.
XX	
PN	EP167182-A2.
XX	
PD	10-JAN-2001.
XX	
XX	
PF	07-JUL-2000; 2000EP-0114090.
XX	
PR	08-JUL-1999; 99JP-0194179.
PR	11-JAN-2000; 2000JP-0118775.
PR	02-MAY-2000; 2000JP-0183766.
XX	
XX	
PA	(HELIX-) HELIX RES INST.

Seq	Sequence	350 AA:
Query Match	99.78%	Score 1757; DB 22; Length 350;
Best Local Similarity	99.78%	Freq. No. 3.2e-164;
Matches 349; Conservative	0; Mismatches 1; Indels 0; Gaps 0	
OY	1 MAVEFVLLALVAGVLGNFSLKSPSSVYFRNGMPTIGERLIPDAALSMGFSVAKEDLSW 60	
Db	1 meafvllalvagvlgnfslkspssvfrngmpipgerlptdaalsmgfsvkedlslw 60	
OY	61 PGLAGNLFRRPATVYMWVGWKLALPGGSVSYPLEANVPFLDSVAANSIHSLFFSEE 120	
Db	61 pglavgnlfrrpatvmwvgnwkallpggsvisyleanvpfldsvaansihslfsee 120	
OY	121 TPVYQLAPSEEVYMWGKASVPEELSVTLRQRLRNLPQENSIVSLSPINSLRRNEVD 180	
Db	121 tpevqlapseevymwgkavseelstvlrlrqrlrnlpqensivslspinslrnmevd 180	
OY	181 LAFUSELDVHLDISLSLRHKILAKDHSPLYSLBLAGLDETCKRGKGRDSQFRDASKTL 240	
Db	181 llfisevlyhdislsllsrhklakdhsplyslblagldetckrgkgrdsqfrdasktl 240	

QY	241	VILQKPFDDKXVSLYSGNMYVVELLYVNSPPTSLIRTRFRIILEAQANPASPYNLAYYN	3000
Db	241	valdjak fadmaysljysgkaveelvyvnsfdtsliifrttllileaqaknaspynlayyn	3000
QY	301	FETSVENMVLIMIALALATTSYXNNMMDGDSIITYRNTQKTRMD	350
Db	301	feysvvfmvltwmlalaavlisyxnnmmdgdyosliiyntmktrktrmd	350

RESULT 4

ID AAB90589 standard; Protein; 350 AA.

AC AAB90589;

DT 01-JUN-2001 (first entry)

DE Human secreted protein, SEQ ID NO: 127.

KM Human; secreted protein; immunomodulatory; antisclerotic;
KM dermatological; antiinflammatory; anti-HIV; cytostatic; cardiac;
KM vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KM neoplastic; anticoagulant; antialzheimers; antiparkinsonian;
KM antimicrobial; vulnerability; vaccine; gene therapy; cancer;
KM protein coordinate data; infection.

OS Homo sapiens.

PN W0200121658-A1

PD 29-MAR-2001

PF 22-SEP-2000; 2000WO-US26013.

PR 24-SEP-1999; 99US-0155709.

PA (HUMA-) HUMAN GENOME SCI INC

PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA,
PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR,
PI Young PE, Wei P, Florence KA;

DR N-PSDB; AAF97929.

PT Nucleic acids encoding 32 human secreted polypeptides, useful for preventing, diagnosing and/or treating e.g. cancers, Parkinson's disease and diabetic retinopathy -

PS Claim 11; Page 809-810; 890pp; English

The present sequence is one of 32 novel human secreted polypeptides. The nucleic acid molecules and polypeptides may be used in the prevention, diagnosis and treatment of diseases such as immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus and human immunodeficiency virus (HIV) infections), hyperproliferative disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases (e.g. Scleritar syndrome, Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic disorders (e.g. corneal graft neovascularisation and diabetic retinopathy), neurological disorders (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease), infectious diseases and/or for promoting wound healing, regeneration and/or chemotaxis. The nucleic acid molecules may be used to produce the secreted polypeptides. They may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples. The polypeptides may be used as antigens in the production of antibodies and in assays to identify modulators of their expression and activity.

\$Q\$ Sequence 350 AA;

50


```

PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154019.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

```

```

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 6.9%; Score 121.5; DB 21; Length 364;
 Best Local Similarity 22.0%; Pred. No. 0.0017;
 Matches 86; Conservative 72; Mismatches 154; Indels 79; Gaps 18;

```

QY 3 VFVVLALVAGVYGNFSEILKPSGVF-----RNGMMPRGGRIPDVAAIS-- 49
DB 6 VFVALLILTLNRYG-----EASGVFLIDGSNGYIRPSSSEALPMSPVEISAASAL 59
QY 50 MGEFVKDLSWPG-----LAVGNLPHRRPRAIVWVWKGYNKLALPGSVISYLENAVP 103
DB 60 IGFPSATLITDAGSSKLKILKPNPFEPRAAFFVLEIAGADMLLETSPSHF--IGNAIR 118
QY 104 FSLDSVANSIHSLFSEETPVVQLAPSEBRVWVGKANSVFEDLSVTLQLRNRL----- 158
DB 119 SSIKS-----DSYKADT-----ELPDNEVVVSVNPSDVTCKDI--NDFASVIGSGYV 166
QY 159 --FOENSVLSSLPILNLSLRNNEVDL-----LFLSELQVYLDHISLSLRKHIAK--DH 207
DB 167 AGAEPSSGLISIPLAG--GANVEFNLEKAEERKFAINLLGYQNIIRGAVSYDDISHGIDR 225
QY 208 SPDIYSLELAGDEITGKRYGDESEQFRDASKILVDALOKFADDMYSLVGNAVVELVTWK 267
DB 226 TAEITVYRFGYIDALAGEYGG--MAKGMVLLSLSKILFNLLETSHG-----QIVGYI 279
QY 268 SFDTSILRIKRTITLFAKQAKNPA-----SPYNLAAYKNFEYSVFFNMVIM--IA 316
DB 280 VLDERVNGESENLIIFGSSIRTSARSMVEVGIPSAALIA-----EVLIVRLTAWLEGIL 334
QY 317 LALAVITSYNTIMNDPQSYIIRMTNOKI 347
DB 335 ILLAETILGYFLIMMPLTKDILLY--SNVKI 363

```

RESULT 8
 AAG11428
 ID AAG11428 standard; Protein: 392 AA.
 XX
 AC AAG11428;
 XX

DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SPO ID NO: 10141.
DE
XX Protein identification: signal transduction pathway; metabolic pathway;
KM hydrolisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138500.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145085.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0148685.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

```

PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

```

Query Match 6.9%; Score 121.5; DB 21; Length 392;

Best Local Similarity 22.0%; Pred. No. 0.0019;

Matches 86; Conservative 72; Mismatches 154; Indels 79; Gaps 18;

```

OY 3 VFVVLATVAGVLGNFSLKPSVVF-----RNGMWPJGGRIDVVALS-- 49
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 34 vfvaallllltnryg-----easvsvffidgsnngylrpsrsealpmysvaisaasal 87
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
OY 50 MGESVKEEDLSWPG-----LAVGNLFRRPRATVWVWVGKVKLALPPGVSIVYLENAPV 103
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 88 lgfapscltladgssklnklknpferpraafvleladgdmlltetspsfaf-lymaair 146
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
OY 104 FSLDSVANSIHSTFSEETPVVLOLAPSEERYVWGKANSVEEDLSVTLQLNRRL----- 158
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 147 sskls-----dsyadct-----elpdnervvsvvsgesdvtckdrl--ndfswlgsasyv 194
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
OY 139 --FOENSVLSSLPLNLSRNNEVDL-----LFLSELQVLHDISSLRKHHLAK--DH 207
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 195 aagapsesglislplag-ganvefnlekeaeerfainlllglyqnlgavsvydlshgldr 253
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
OY 208 SPDLYSELTAGLDEGKRYGDSQFPAASKTLIVDALOKFADWYSLYGGMVVELVTVK 267
    ||||| | | | | | | | | | | | | | | | | | | | | | | |

```

```

DB 254 taeltvtrfgjidalageyqg--makgmdvllstlsklfnlletsuky----qlvgyl 307
OY 268 SFTSILRKRTTILEAKQAKNPA-----SPYMLAKTYFEYSVFMVLMIM-IA 316
    | : : : | : : : | : : : | : : : | : : : | : : : |
DB 308 vlderegesenllnfgsrsarsmvevegijsaaila-----evllvrltlawllgll 362
OY 317 LALAVIDTSYNIMWMDPGYSITYRMTNOKI 347
    | : | : | : | : | : | : | : | : | : | : | : | : |
DB 363 llaatllyyflmmptlktllyl--snvkl 391

```

RESULT 9

AAW98378 standard; Protein; 428 AA.

AAW98378;

31-MAR-1999 (first entry)

H. pylori GHPO 1375 protein.

GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;

peptic ulcer disease.

Helicobacter pylori.

W09843478-A1.

08-OCT-1998.

01-APR-1998; 98WO-US06371.

29-JUL-1997; 97US-0902615.

01-APR-1997; 97US-0833457.

24-JUN-1997; 97US-0881227.

(HUMA-) HUMAN GENOME SCI INC.

(INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;

WPI: 1998-542293/46.

N-PSDB; AAX14097.

Claim 8; Page 581-583; 2054pp; English.

This sequence represents a Helicobacter pylori GHPO protein of the invention. The polypeptides can be used for preventing or treating Helicobacter infections, and gastroduodenal diseases associated with these infections, including acute, chronic, and atrophic gastritis, and peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used for the production of antibodies. The products can also be used for detection and diagnosis.

Sequence 428 AA:

Query Match 5.9%; Score 103.5; DB 19; Length 428;

Best Local Similarity 22.9%; Pred. No. 0.13;

Matches 72; Conservative 35; Mismatches 115; Indels 93; Gaps 15;

```

OY 93 VISVPLENA-----VPSLDSVANSIHSTFSEETPVVLOLAPSEERY-----YMG 138
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 21 viskelinaangvkvrtllldt--ngldsdtsd-----imlfnfnknevkifmppyira 73
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
OY 139 KANSVE---DLSTVTLROLNRRLFOENSVLSSLPLNLSRN--NEVDLFLSELQVL-- 190
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 74 kglyrlemadyerlkkrmknkflfvdnfavnlgvgrnlgndyfdndldntfl-dlaalff 132
    ||||| | | | | | | | | | | | | | | | | | | | | | | |

```



```
OY 191 -----HDIS-SLSRKHAKADHSPDLYSLEAGLDFTGKRYGDESE 231
      ||| ||| : ||| |||
Db 133 gvasakakesferwyrfhnsrpsipvllrthrkln-----akeiaiknekipisaekdn 186
      ||| ||| : ||| |||
OY 232 QFRDASITLVDALOFADDMYSLYGNAYVELYFWKSPDTSLIRKPTTILEKOKNPAS 291
      ||| ||| : ||| |||
Db 187 qfekvndfidrtdy---qypllygnafildspkldpplypklafekalknkd 242
      ||| ||| : ||| |||
OY 292 PYNLAYV-----NEFSVFNWLMITALALAVITSYNINMDPG 334
      ||| ||| : ||| |||
Db 243 svfiassyfipgkkmkikfnqiskgjeimltn-----slsldaiyygawe----- 291
      ||| ||| : ||| |||
OY 335 YDSIYRMTNOKIRM 349
      ||| ||| : ||| |||
Db 292 -----tyrnlqvlrm 300
      ||| ||| : ||| |||

RESULT 10
AAG11430
ID AAG11430 standard; Protein; 318 AA.
XX
AC AAG11430;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10143.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
```

PR	14-AUG-1999	9905-0147302
PR	05-AUG-1999	9905-0147132
PR	05-AUG-1999	9905-0147360
PR	06-AUG-1999	9905-0147303
PR	06-AUG-1999	9905-0147416
PR	09-AUG-1999	9905-0147433
PR	09-AUG-1999	9905-0147435
PR	10-AUG-1999	9905-0148111
PR	11-AUG-1999	9905-0148319
PR	12-AUG-1999	9905-0148341
PR	13-AUG-1999	9905-0148455
PR	13-AUG-1999	9905-0148664
PR	16-AUG-1999	9905-0149368
PR	17-AUG-1999	9905-0149175
PR	18-AUG-1999	9905-0149426
PR	20-AUG-1999	9905-0149722
PR	20-AUG-1999	9905-0149723
PR	20-AUG-1999	9905-0149929
PR	23-AUG-1999	9905-0149930
PR	25-AUG-1999	9905-0149930
PR	26-AUG-1999	9905-0150566
PR	27-AUG-1999	9905-0150864
PR	27-AUG-1999	9905-0151066
PR	27-AUG-1999	9905-0151060
PR	28-AUG-1999	9905-0151303
PR	31-AUG-1999	9905-0151348
PR	30-AUG-1999	9905-0151438
PR	01-SEP-1999	9905-0151963
PR	07-SEP-1999	9905-0152363
PR	10-SEP-1999	9905-0153070
PR	13-SEP-1999	9905-0153758
PR	15-SEP-1999	9905-0154039
PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154779
PR	22-SEP-1999	9905-0155139
PR	23-SEP-1999	9905-0155386
PR	24-SEP-1999	9905-0155559
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156596
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159294
PR	13-OCT-1999	9905-0159295
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159330
PR	14-OCT-1999	9905-0159637
PR	14-OCT-1999	9905-0159637
PR	14-OCT-1999	9905-0159637
PR	18-OCT-1999	9905-0159684
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160981
PR	25-OCT-1999	9905-0161404
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161359
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161920
PR	28-OCT-1999	9905-0161992
PR	29-OCT-1999	9905-0162142

Query Match Summary:

Local Similarity	21.9%	Pred.	No. 0.13:	Score	101.5;	Df	21;	Length	318;
Matches	75;	Conservative	66;	Mismatches	141;	Indels	61;	Gaps	16;

Sequence Alignment Details:

```
OY      38 PEERIPVVAALSKGFSVKEDLSMP-----LAVGMLEFRRPRATVMWKKGVNKLALPPG   91  
       | :| :|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     3 pveisaavsal- lglfaaaclladgasklnkllkpnrferpraa fvelagaddmllets    61  
  
OY     92 SVTSPYLEENAPFPSLDVSANSHSLSEETRPVVYLQLPSEEKRYMVGKANVFDELSVTL   151  
       ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     62 pshef-f-lgnatirsiks-----dsykadt-----elpdnevvvsnepsdvrdkd--  108  
  
OY     152 RLGNRL-----FOENVSLSTPLNSLNNNVDL-----LFTELQVLHDISLL    197  
       ::::| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     109 ndfaswjggsgyaagaessqlslrpaq-gaanefnlekeaeerftalnlgynlrfqay    167  
  
OY     198 SRHKHLAK--DHSDPLYSELGLADEIGKRKYDSEDGFADASKILLDALOKPADMYSLX    255  
       |:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     168 syvdldishgidrlaelvrfggdalayaegyg9-makgmadvllstlsklfnllesth    225  
  
OY     256 GSNVNVELVYSFPSTSIRKTTLTELEAKOKKNPA-----SPYNLAAYKYNPEYSV    305  
       :::| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     226 ky----giygvavidervngesenllnfgsrrsarasmwervegilpsaalaaaa---evil 276  
  
OY     306 VENNWLIM-LALALAIVITSYNMWNMDPGDISIIRMTNKI    347  
       | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
Db     277 vrtlawltgilliatllgyvlmmpblktclly--snvk1    317  
  
RESULT  11  
AAV10994  
ID AAV10994 standard; Protein; 412 AA.  
XX  
AC AAV10994;  
DX DT  
DY 08-JUN-1999 (first entry)  
DE H. pylori ORF 09cel0413_35336707_f2_9 secreted protein.  
DM XX  
KW Vaccine; probe; diagnostic; ORF; cell envelope protein;  
KM secreted protein; cellular protein.  
KS XX  
OS Helicobacter pylori.  
PS OS  
PN PN  
FN W09J1B323-A1.  
PD PD  
PY 07-MAY-1998.  
PX XX  
PF PF  
PI 28-OCT-1997; 97WO-US19575.  
PT PR  
PT 14-JUL-1997; 97US-OB91928  
PR 28-OCT-1996; 96US-0739150  
PR 06-DEC-1996; 96DS-0759739.  
RX XX  
XA (ASTR ) ASTRA AB.  
XX XX  
XP Alm RA, Smith D;  
PI PI  
DR MPI: 1998-271811/24.  
DR N-PSSDB; AAX30461.  
DT DR  
PT PT  
PT Products for the detection, prevention and treatment of H. pylori  
PT Infections  
XT XX  
XX Claims 27, 31; Page 201-202; 279pp; English.  
CC Recombinant or substantially pure preparations of H. pylori polypeptides  
CC are disclosed, together with the nucleic acids encoding them. In all,  
CC 73 ORFs are shown. The proteins are variously cell envelope proteins,  
CC secreted proteins or other cellular proteins. Vaccines containing the  
CC nucleic acids or proteins are claimed, as are probes containing at least
```

CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful
 CC for treating or reducing the risk of H. pylori infections, and the
 CC probes can be used diagnostically for detecting the presence of
 CC Helicobacter in a sample. The products are also of use in screening
 CC for compounds having the ability to interfere with the H. pylori life
 CC cycle or to inhibit H. pylori infection.

XX Sequence 412 AA;

Query Match 5.7%; Score 101; DB 19; Length 412;
 Best Local Similarity 20.8%; Pred. No. 0.21;
 Matches 69; Conservative 50; Mismatches 93; Indels 120; Gaps 17;

QY 48 LSMGFVYKEDLSMPGLAVGNLFH-----RPRATVAVWVKGYNKIALPPGCVI 94
 DB 97 msjlsqkvalngkkltskminleqkkllelktkqqlvlnlmngle----- 146
 QY 95 SYPLENAVPE--SLDSVANSIRSLFSEETPVYQLAPSE--ERYVWVGKANSVFEDLSV 149
 DB 147 nyknqgelelntalknlenllyqanhsspdllailelkslelqk-----ndlev 201
 QY 150 TLR---QLRNLFOENSVLSPLNSRNNEVDLLSELQVLDHDISSLSRKHIAK 205
 DB 202 alssshysmgelttkeneilsiapkn-fefme-----gelmlisa----- 241
 QY 206 DHSFDLYSLEAGLDE-----IGKRYGDSF--QFRDASKILVDALOKFADDM 251
 DB 242 ----tmdlaarldeekagkdtllakksfledinvlgyvfr-----skgyynydm 289
 QY 252 YS-----LVGNAVVELVTVVTSFDT-----SLIRKTRIT--L 281
 DB 290 fsvslslprip1y9kqa--klveqgkkeslaifsevenakntrlaklkletlqnl 347
 QY 282 EA-----KQAKNPASPVNLAKYKFEYSVFN 308
 DB 348 eslnklkqnekiaqlayldlktngdynayn 379

RESULT 12

AA039631
 ID AAR39631 standard; Protein: 1103 AA.

AC AAR39631;

DT 16-DEC-1993 (first entry)

DE Neurofibromatosis type 1 polypeptide.

KW Non-defective gene; NF-1; treatment; tumours; human; detection; ss.

OS Homo sapiens.

PN US5227292-A.

PD 13-JUL-1993.

PF 12-JUL-1990; 9005-0551531.

PR 12-JUL-1990; 9005-0551531.

PA (UTAH) UNIV UTAH.

PI Cawthon RM, O'Connell P, Viskochil DH, White RL;

DR WPI: 1993-235118/29.

DR N-PSDB; AAQ46263.

PT cDNA encoding neurofibromatosis type 1 gene - for detecting

PT defective NF1 genes and tumours caused by such genes

PS Disclosure; Fig 7; 59pp; English.

CC The sequence is that of the neurofibromatosis type 1 (NF1)
 CC polypeptide which may be used therapeutically in the treatment of
 CC diseases associated with defective NF1 genes, e.g. tumours.

XX Sequence 1103 AA;

Query Match 5.7%; Score 100.5; DB 14; Length 1103;
 Best Local Similarity 22.2%; Pred. No. 1.1; Indels 101; Gaps 18;
 Matches 86; Conservative 60; Mismatches 140;

QY 13 GVLGNEPSILKSPGVSFVRGNGMPDGE-----RIPDYAALSMGFVYKEDLSMPGLAV 65
 DB 416 g9lgs1kkaemadavlaasgnvklvskvlgmcklkdkcslspptleghlmdl 475
 QY 66 GNLFRPRATVAVWVKGYNKIALPPG-----SVISYPLENAVPELSDS-----VANSTHS 115
 DB 476 -----larymlmstfmsldvaahpylfnvltf-lvatpislraathglvlnlshs 527
 QY 116 L-----FSEETPVYQLAPSE--ERYVWVGKANSVFEDLSVTLR-OLRNLFOENSV 164
 DB 528 lctcsqhlstseetkqvlrlstetslpktyllfjiskvksaavlatrsydrstspgsv 587
 QY 165 -LSLPLNSLRNNEVDLLFSELQVLDHDISSL--LSRKHIAK-----DHSFDLYSLELA 217
 DB 588 erefalsletvtea-llelme-acmrldptckwldqwtelagrfafgynpslqpraly 645
 QY 218 GLDEIGKR--YGEDEQFRDASKILVDALOKFADDMYS----- 253
 DB 646 vfgclskrvshgikqlrlrlskalescl- gpdtyngsvlleetvialctklp1lndk 703
 QY 254 -----LVGNAVVEL--VTVKSFDTSLIRKTRITLEAKQAKNPASP----- 292
 DB 704 splkhalfwavavqldevnlysaqtaileqnlhldslrlfndkspeevfmaltrple 763
 QY 293 -----YNLAYKRYKFEYSVFN 310
 DB 764 whckgmthfvglnfnfnfalvghll 790

RESULT 13

AA039921
 ID AAR59921 standard; Protein: 2485 AA.

AC AAR59921;

DT 22-FEB-1995 (first entry)

DE RAS associated GAP NF201.

KW Ras; GTPase activating protein; GAP; GAP related domain; GRD;

KW pK10; pK11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;

KW neurofibromatosis type 1; NF1.

OS Homo sapiens.

PN W09416069-A.

PD 21-JUL-1994.

PF 12-JAN-1994; 94WO-US00198.

PR 15-JAN-1993; 93US-0004824.

PA (SCHE) SCHERING CORP.

PI Kaziro Y, Nakafuku M;

DR WPI: 1994-249216/30.

PT Blocking Ras-induced effects on a cell - by introducing a GTPase
 PT activating protein to the cell, used esp. in treatment of cancers

PS Disclosure; Page 36-44; 87pp; English.

XX Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD) was
 CC cloned into the yeast expression vector pK10 to obtain pK11. The
 CC pK11 DNA was mutagenized by hydroxylamine in vitro and transformed
 CC into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type
 CC RAS2Val19 mutation. The heat shock sensitivity of the clones was
 CC checked. Plasmid DNAs were recovered, re-transformed into TK161-
 CC R2V-D, and phenotypic reversion was examined. 2 Clones, NF201 and
 CC NF204 (given in AAR59922), which had strong suppression activity for
 CC RAS2Val19, were selected. The mutant NF1-GRDs were also able to
 CC inhibit v-Ras-induced transformation in mammalian cells.

XX Sequence 2485 AA;

Query Match 5.7%; Score 100.5; DB 15; Length 2485;

Best Local Similarity 22.2%; Pred. No. 3.9;

Matches 86; Conservative 60; Mismatches 140; Indels 101; Gaps 18;

QY 13 GVLGNEFSLKSPGVSFRNGNMPDGE-----RIPDVALLSMGFSYKEDLSMPGLAV 65
 DB 1667 gglgslkaevmadtvalaasgnvklvskvigrmckltdkclspptleqhlmwddai 1726
 QY 66 GNLFHRRATVWVWKGYNKALPPG-----SVISYPLENAVPSLDS-----VANSIHS 115
 DB 1727 -----larylmisfnslsdaahpylflhvvtf-lvatgplslraetghlvinihs 1778
 QY 116 L-----FSEETPVVQLAPSE---ERYVMGKANSYFEDLSVTLR-QLNRRLFOENSV 164
 DB 1779 lctcsqhlhseetkqvlrlstelsjpktyllfglskksaavaftrsfydrstspgsy 1838
 QY 165 -LSSLPLNLSLRNNEVDLFLSELYVLDHISL--LSRHKHLAK---DHSDDLSESLA 217
 DB 1839 erefaltstleevtea-lleime-acmrdipckwldqwtelagratfagynpslgrpralv 1896
 QY 218 GLDDEIGR--YGEDSEQFRDASKIYDALQKPADMYS----- 253
 DB 1897 vfgclskrvshqqlkqllrlskalescll--gpdctynsqvllaeatvaltklqlinkd 1954
 QY 254 -----LYGNAVVEL--VTVKSFDTSILRKTITILEAKOKNPASP----- 292
 DB 1955 splhkalfwavavqldevnlysgatlllegnlhtldslrlfndkspeevfmairple 2014
 QY 293 -----YNLAYKYNFEYSVFNWV 310
 DB 2015 whckgmhfvglfnfnfnalvaghl 2041

RESULT 14
 AAR59922
 ID AAR59922 standard; protein; 2485 AA.

XX AAR59922;

XX 22-FEB-1995 (first entry)

XX RAS associated GAP NF204.

XX Ras: GTPase activating protein; GAP: GAP related domain; GRD:
 KM pK10; pK11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;
 KW neurofibromatosis type 1; NF1.

XX Homo sapiens.

XX OS

XX PN W09416069-A.

XX PD 21-JUL-1994.

XX PE 12-JAN-1994; 94WO-US00198.

XX PR 15-JAN-1993; 93US-0004824.

PA (SCHE) SCHERING CORP.

PI Kaziro Y, Nakafuku M;

XX WPI; 1994-249216/30.

XX Blocking Ras-induced effects on a cell - by introducing a GTPase
 PT activating protein to the cell, used esp. in treatment of cancers
 PS Disclosure; Page 44-52; 87pp; English.

CC Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD) was
 CC cloned into the yeast expression vector pK10 to obtain pK11. The
 CC pK11 DNA was mutagenized by hydroxylamine in vitro and transformed
 CC into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type
 CC RAS2Val19 mutation. The heat shock sensitivity of the clones was
 CC checked. Plasmid DNAs were recovered, re-transformed into TK161-
 CC R2V-D, and phenotypic reversion was examined. 2 Clones, NF201
 CC (given in AAR59921) and NF204, which had strong suppression activity
 CC for RAS2Val19, were selected. The mutant NF1-GRDs were also able
 CC to inhibit v-Ras-induced transformation in mammalian cells.

XX Sequence 2485 AA;

Query Match 5.7%; Score 100.5; DB 15; Length 2485;

Best Local Similarity 22.2%; Pred. No. 3.9;

Matches 86; Conservative 60; Mismatches 140; Indels 101; Gaps 18;

QY 13 GVLGNEFSLKSPGVSFRNGNMPDGE-----RIPDVALLSMGFSYKEDLSMPGLAV 65
 DB 1667 gglgslkaevmadtvalaasgnvklvskvigrmckltdkclspptleqhlmwddai 1726
 QY 66 GNLFHRRATVWVWKGYNKALPPG-----SVISYPLENAVPSLDS-----VANSIHS 115
 DB 1727 -----larylmisfnslsdaahpylflhvvtf-lvatgplslraetghlvinihs 1778
 QY 116 L-----FSEETPVVQLAPSE---ERYVMGKANSYFEDLSVTLR-QLNRRLFOENSV 164
 DB 1779 lctcsqhlhseetkqvlrlstelsjpktyllfglskksaavaftrsfydrstspgsy 1838
 QY 165 -LSSLPLNLSLRNNEVDLFLSELYVLDHISL--LSRHKHLAK---DHSDDLSESLA 217
 DB 1839 erefaltstleevtea-lleime-acmrdipckwldqwtelagratfagynpslgrpralv 1896
 QY 218 GLDDEIGR--YGEDSEQFRDASKIYDALQKPADMYS----- 253
 DB 1897 vfgclskrvshqqlkqllrlskalescll--gpdctynsqvllaeatvaltklqlinkd 1954
 QY 254 -----LYGNAVVEL--VTVKSFDTSILRKTITILEAKOKNPASP----- 292
 DB 1955 splhkalfwavavqldevnlysgatlllegnlhtldslrlfndkspeevfmairple 2014
 QY 293 -----YNLAYKYNFEYSVFNWV 310
 DB 2015 whckgmhfvglfnfnfnalvaghl 2041

RESULT 15
 AAR22268

ID AAR22268 standard; protein; 2818 AA.

XX AAR22268;

XX 06-MAY-1992 (first entry)

XX Nf1 gene product.

XX von Recklinghausen neurofibromatosis disease; autosomal dominant;
 KM gene therapy.

XX Homo sapiens.

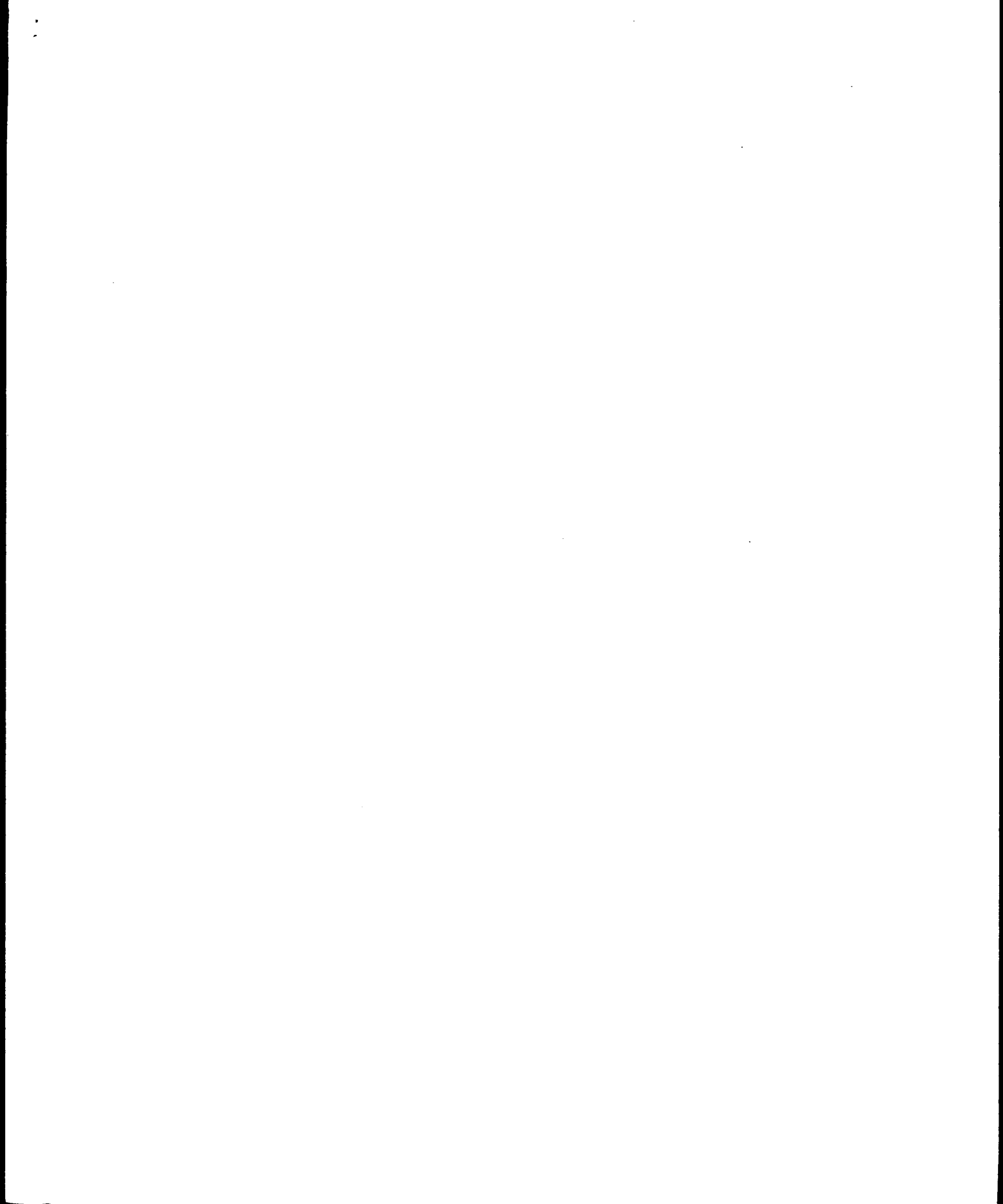
XX OS

XX	MO9200387-A.
PN	
XX	09-JAN-1992.
PD	
XX	
XX	28-JUN-1991; 91MO-US04624.
XX	
PR	29-JUN-1990; 90US-0547090.
PA	
XX	(UNMI) UNIV OF MICHIGAN.
XX	
PI	Collins FS, Wallace MR, Marchuk DA, Andersen LB, Gutmann DH;
XX	
DR	WPI; 1992-041568/05.
DR	N-PSDB; AAQ20602.
XX	
PT	DNA sequences to von-Recklinghausen neurofibromatosis gene - and
PT	derived amino acid sequences and probes for screening NFI in early
PT	stages of disease
XX	
XX	
PS	Claim 25; Page 67; 122pp; English.
XX	
CC	This is the amino acid sequence of the von Recklinghausen neuro-
CC	fibromatosis (NFI) gene product. It and antibodies raised to it
CC	can be used in hybridisation and immunological assays to screen for
CC	the presence of a normal or defective NFI gene product. Functional
CC	assays to measure levels of gene function can also be used for
CC	diagnosis or to monitor treatment. Patient therapy through
CC	supplementation with the normal NFI product which can be
CC	produced by recombinant techniques is also possible.
XX	
50	Sequence 2818 AA;

Query Match	5.7%	Score 100.5;	DB 13;	Length 2818;
Best Local Similarity	22.2%;	Pred. No. 4.8;		
Matches	86;	Conservative	60;	Mismatches 140;
				Indels 101;
				Gaps 18;

[illegible]

Search completed: July 3, 2001, 16:18:21
Job time: 168 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:16:43 ; Search time 16.08 seconds

(without alignments)
1638.028 Million cell updates/sec

Title: US-09-284-320-6

Perfect score: 1763

Sequence: 1 MAVFVVLALVAGVAGNEFS.....MDPGYDSTIYRMTNOKIRMD 350

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1667	94.6	330	2 T08667	hypothetical prote
2	192.5	10.9	324	2 T23876	hypothetical prote
3	121.5	6.9	364	2 T52336	probable type 1 me
4	114	6.5	663	2 T40039	hypothetical prote
5	109.5	6.2	408	2 T25243	hypothetical prote
6	105	6.0	450	2 S35732	data protein - Sp1
7	105	6.0	458	2 T19941	hypothetical prote
8	104	5.9	412	2 G64685	hypothetical prote
9	103.5	5.9	502	2 F64543	conserved hypotet
10	103.5	5.9	732	2 T43619	secreted protein k
11	103.5	5.9	732	2 S30060	protein kinase ypk
12	103.5	5.9	925	2 C84538	probable LRR recep
13	103	5.8	470	2 A71685	guanosine pentaplo
14	103	5.8	979	2 J00894	P115 protein - Myc
15	102.5	5.8	1004	2 A39611	probable GTP-bindi
16	101	5.7	412	2 B71831	hypothetical prote
17	101	5.7	689	2 T40364	hypothetical prote
18	100.5	5.7	1358	2 A29360	S1R4 protein - Yea
19	100.5	5.7	2818	2 B55282	neurofibromin I -
20	100.5	5.7	2820	2 J05196	neurofibromin I -
21	99	5.6	524	2 S77851	probable aspartate
22	98.5	5.6	377	2 S21302	succinate dehydrog
23	98.5	5.6	471	2 T48743	probable 26S ATP/u
24	98.5	5.6	1038	1 MMRRCB	myosin beta heavy
25	98.5	5.6	1935	1 A37102	myosin beta heavy
26	98	5.6	561	2 F75191	hypothetical prote
27	98	5.6	864	2 T08575	protein kinase hom
28	97.5	5.5	758	2 H75013	hypothetical prote
29	97.5	5.5	884	2 S73302	preprotein transio

30	97.5	5.5	2825	2	154352	neurofibromin - mo
31	96.5	5.5	533	2	G72593	hypothetical prote
32	96.5	5.5	574	2	S17199	CtBP1-like protein
33	96.5	5.5	1062	2	F83335	RND multidrug effl
34	96.5	5.5	1062	2	T30830	hypothetical prote
35	96.5	5.5	1934	2	I48153	myosin heavy chain
36	96.5	5.5	1938	1	S06005	myosin alpha heavy
37	96	5.4	390	2	E81122	bacteriophage DNA
38	95.5	5.4	1020	2	B82427	sensor protein Tor
39	95.5	5.4	1209	2	T21455	hypothetical prote
40	95.5	5.4	2471	2	T42977	large tegument pro
41	95	5.4	2253	2	T30336	nuclear/mitotic ap
42	94.5	5.4	1935	1	S06006	myosin beta heavy
43	94.5	5.4	1935	2	A59286	myosin heavy chain
44	94	5.3	1778	2	T50074	probable nucleopor
45	93.5	5.3	818	2	T31464	stage II sporulati

ALIGNMENTS

RESULT 1
T08667
hypothetical protein DKFZp54700510.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
C:Accession: T08667
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: 216466
A:Accession: T08667
A:Molecule type: mRNA
A:Residues: 1-330 <POU>
A:Cross-references: EMBL:AL049929
A:Experimental source: fetal brain; clone DKFZp54700510
C:Genetics:
A:Note: DKFZp54700510.1

Query Match 94.6%; Score 1667; DB 2; Length 330;

Best Local Similarity 99.7%; Pred. No. 1.6e-114;

Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	21	ILKSPGVFRNGNMP	IPGERIPDVAALSMGFSYKEDLSMPGLAVGNLFHRRPRTAVWV	80
DB	1	ILKSPGVFRNGNMP	IPGERIPDVAALSMGFSYKEDLSMPGLAVGNLFHRRPRTAVWV	60
QY	81	KGVKALALPPGVSIV	PLENNAVPPSLDSVANSIHSLFSEETPVVQLALPSEERYWYWGKA	140
DB	61	KGVKALALPPGVSIV	PLENNAVPPSLDSVANSIHSLFSEETPVVQLALPSEERYWYWGKA	120
QY	141	NSVFEDLSVTLKQL	ENRRLFOENSVLSPLNSLSRNNNEVDLFLSELQVLDHISLSLRH	200
DB	121	NSVFEDLSVTLKQL	ENRRLFOENSVLSPLNSLSRNNNEVDLFLSELQVLDHISLSLRH	180
QY	201	KHLAKHSDPLSL	SLDLDELIGKRGEDSEDFRDASKIIVDALOKFADDMYSLYGNAV	260
DB	181	KHLAKHSDPLSL	SLDLDELIGKRGEDSEDFRDASKIIVDALOKFADDMYSLYGNAV	240
QY	261	VELVYKSPDTSIL	IRKRTTILEAKQAKNPASPYNLAYKYNREXSVVENVYMTATATA	320
DB	241	VELVYKSPDTSIL	IRKRTTILEAKQAKNPASPYNLAYKYNREXSVVENVYMTATATA	300
QY	321	VITSYVINNM	DPGYSITTYRMTNOKIRMD	350
DB	301	VITSYVINNM	DPGYSITTYRMTNOKIRMD	330

RESULT 2
T23876
hypothetical protein R03EL.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T23876
R:McMurray, A.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z19812
A:Accession: T23876
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-324 <WTL>
A:Cross-references: EMBL:592837; PIDN:CAB07401.1; GSPDB:GN00028; CESP:R03E1.2
A:Experimental source: clone R03E1
C:Genetics:
A:Gene: CESP:R03E1.2
A:Map position: X
A:Introns: 59/1, 180/2, 293/3

Query Match	10.9%;	Score 192.5;	DB 2;	Length 324;
Best Local Similarity	22.3%;	Pred. No. 9.6e-07;		
Matches 86;	Conservative 64;	Mismatches 134;	Indels 101;	Gaps 13

[illegible]

```

QY 115 SLFSEEPY-----VLQAPSEERYWYGKANSVFEDLSVTLROLNRLFQENSVL 166
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 112 GADRENQVSAGITGSMALAAQEA-----DATV-ITKTRSTL----- 151

```

0y 167 SLPLNSLSRNNEVDLLFLSELQVLHDLSLSLRHKHLAKHSDPLYSLELAGLDEIGKRY 226
 : || : : : : : | : : : : : |
 Db 152 -----YKFLFAVYOLAAL-KSAGVOMNNADVEPSTGCV----- 199

Qy 227 GEDSEQFRDASKILVDALOKFADDMYSLYGNVVELY-----T 265
:| | : |:: : ||| |::|:

QY 266 VKSFDTSIRKTRTILEAQAKNPASPYNLAKYKNFEYSVENMWLWIMIALALAVIITS 325

DU 243 LGKUCJNSNFIQLEARENQYIVPSS-----DIPATFAIFLGLVILVVALIYIV 299
 QY 326 YNIWMNDPGYDSIIYRMTNQKIRMD 350

Dd 300 VGMASIDPEKDSIYYRMTTTRMKD 324

RESULT 3
T52336
Probable type I membrane protein [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C:Accession: T52336
R:COO, J.H.: Park, A.R.: Park, W.J.: Park, O.K.
Plant Mol. Biol. 41, 415-423. 1999

A: Reference number: 225437; M0ID:20064977
 A: Accession: T52336
 A: Accession: T52336

A;Molecule type: mRNA
A;Residues: 1-364 <GO>
A;Cross-references: EMBL:AF104329; PIDN:AAD11797.1

A;Gene: PMH

```

Query Match      6.9%; Score 121.5; DB 2; Length 364;
Best Local Similarity 22.0%; Pred. No. 0.18;
Matches 86; Conservative 72; Mismatches 154; Indels 79; Gaps 18;

3 VFVVLLAVAGVLGNEFSLKPGSVVF-----RNCNNPDPGEITPDVAALS- 49

```

[illegible]

RESULT 4
T40039

C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C:Accession: TA0020, TA0704

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21901

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-663 (W003)

A: Experimental source: strain 972h⁻; cosmid c28E12
R: Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
A: Cross-references: EMBL:AL031511; PIDN:CAA20649.1; GSPDB:GN00067; SPDB:SP

A:Reference number: Z21875
A:Accession: T40794
A:Status: preliminary; translated from GR/EMBL/DBIT

A:Cross-references: EMBL:AL049769; PIDN:CAB42374.1; GSPDB:GN00067; SPDB:SP

C:Genetics:
A:Gene: SPBC28E12.02; SPAC9B6.13
A:Map position: 2

Query Match 6.58; Score 114; DB 2; Length 663;

Matches 62; Conservative 43; Mismatches 92; Indels 92; Gaps

```

Db      414  CVSSSELIVSTGIVEING-----IGEK-----MSFPLKLISIP-----451

```

Db 452 PTEFAQAIAITCAGSVEML-LKTINGIEYFGQENTVPIADMDKASKIFYKGGSGWHQILL 510

```

511 EAPFDQDFISCKKNGKLDKVKQCCFNLKNG-----DILFCP 548

```


QY 166 ELQVLDHDSLSLRKHLAKDHPDLYSLAGL-----DEIGKR 225
 Db 549 QSTISFTV-----DIYSDLEERYTKGKNTMLLEFPAENHGYVPELHKK 592
 QY 226 -YGEDSEQFQDASKILVDLQKPADMYSLVGNNAVVELTVKVSFDTSL 273
 Db 593 LIGRGEQIQRYTK-LVNSYIESTTPPCYGHNVLLR--TPKSFSENL 638

RESULT 5

T25243
 hypothetical protein T24D5.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
 C:Accession: T25243
 R:Wilkinson, J.
 submitted to the EMBL Data Library, November 1995
 A:Reference number: Z20003
 A:Accession: T25243
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-408 <M11>
 A:Cross-references: EMBL:Z68012; PIDN:CAA92020.1; GSPDB:GN00028; CESP:T24D5.2
 A:Experimental source: clone T24D5
 C:Genetics:
 A:Gene: CESP:T24D5.2
 A:Map position: X
 A:Introns: 114/2; 134/3; 206/3; 309/3; 355/2
 C:Superfamily: Caenorhabditis elegans hypothetical protein T24D5.2

Query Match 6.2%; Score 109.5; DB 2; Length 408;
 Best Local Similarity 20.8%; Pred. No. 1.6;
 Matches 43; Conservative 47; Mismatches 84; Indels 33; Gaps 7;
 QY 81 KGVRK-LALPPGSVIS-----YPLENAVPPSLDSVANSIHSI-L-FSEETPVYLQAPSEE 132
 Db 103 KAMNQFLKGPGEVLAKYFQCHLKTMYKKNYTRLDAYKMFSPRDSIILIRSDGS 162
 QY 133 RYVWVGKANSVFEDLSVTLROLRRLFOENSYSLSPLNSRNNEDLLFSELQVYLD 192
 Db 163 TVHNGISNLTFTNRMALIEELRNMISQNTIRYKSLFVTKNSQO-----KGEMLDKC 215
 QY 193 ISSLSRHKHLAKDHPDLYSLAGLDEIGK-----RYGE--DSEQFQDASKILVDA 243
 Db 216 VSTTALNLTAVVWDSPPNITELVTLHRPMLNKTFLMCGSGEYGNIMELEQYRNAMNLKDG 275
 QY 244 LQKPADMYSLVGNNAVVELTVKVSFDTSL 270
 Db 276 LSKITDQ-----LNSITATNFE 293

RESULT 6

S35732
 dnaa protein - Spiroplasma citri
 C:Species: Spiroplasma citri
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 02-Feb-2001
 C:Accession: S35732
 R:Ye, F.; Taigret, F.; Bove, J.
 submitted to the EMBL Data Library, December 1992
 A:Description: Nucleotide sequence and genetic organization at the replication origin (C
 A:Reference number: S35732
 A:Accession: S35732
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-450 <YEP>
 A:Cross-references: EMBL:219108; NID:g49345; PIDN:CAA79521.1; PID:g49346
 C:Genetics:
 A:Genetic code: SGCC3
 C:Superfamily: replication initiation protein dnaa
 C:Keywords: DNA binding; nucleotide binding; P-loop
 F146-153/Region: nucleotide-binding motif A (P-loop)
 F1208-212/Region: nucleotide-binding motif B

Query Match 6.0%; Score 105; DB 2; Length 450;
 Best Local Similarity 20.4%; Pred. No. 3.9;
 Matches 67; Conservative 41; Mismatches 115; Indels 106; Gaps 13;

QY 81 KGVRK-LALPPGSVIS-----YPLENAVPPSLDSVANSIHSI-LFSEETPVYL 124
 Db 83 KNINKOA-----SVISKIDLTENNNAIAYENTTFENRYR-----GDSNHEKMOALAVA 132
 QY 125 LQAPSEERYWVGKANSVFEDLSVTLROLRRLFOENS-----VLSSIPLNSLRN 176
 Db 133 LDLCCKMNPFLFYIGDSGLGKTHLHALENKVEIKYTNRRKYKAKADEFGKIAMDIINOG 192
 QY 177 NEV-----DLFLSEIOLV-----HDISSLSRKH--LAKDHPDLY 212
 Db 193 HEITFAKTSYDITDCLLDIDQLLAKRKKTNELFFHIFNSYLEKKQIVITSDKYPD-- 250
 QY 213 SLELAGD-----EIGKRYGEDSEQFQDASKILVDAQ-----KF 247
 Db 251 --DLGCFEARIISRFSYGLSIGLSDPDEFTALKLEOKLKHONNLGLFSEESLEFIALNF 308
 QY 248 ADDMYSLYGG-----NAVELATV--KSPDTSLIKRTTILEAKQAKNPASP 292
 Db 309 NSDYRKLEGAIRKRLFLAVNMKRNKNETILLADYEKAFKNAFLONNKKITPKIKQIYADS 368
 QY 293 YNLAKYKNEYSVFNMVLMITALLALAV 321
 Db 369 YNITIKAMSKSRVSN---VMOAROLAM 393

RESULT 7

T19941
 hypothetical protein C44H4.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T19941
 R:Smey, R.
 submitted to the EMBL Data Library, August 1996
 A:Reference number: Z19200
 A:Accession: T19941
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-458 <M12>
 A:Cross-references: EMBL:Z79598; PIDN:CAB01867.1; GSPDB:GN00028; CESP:C44H4.1
 A:Experimental source: clone C44H4
 C:Genetics:
 A:Gene: CESP:C44H4.1
 A:Map position: X
 A:Introns: 24/3; 157/1; 257/3; 372/3; 425/1

Query Match 6.0%; Score 105; DB 2; Length 458;
 Best Local Similarity 22.5%; Pred. No. 4;
 Matches 69; Conservative 47; Mismatches 116; Indels 74; Gaps 12;

QY 18 EFTSLKPGSVVRNKNMPPGERIPVYALSMGCSYKEDLSWGLAVCNLHRRRAVYM 77
 Db 149 QKLTSLPDLI-----EYVDNAFLS-----THSDSLK 179
 QY 78 VWVGVKVLALPPGSVISYPLENAVPPSLD-SVANSIHSI-LFSEETPVYLQAPSEERYM 136
 Db 180 LDLSANNMLTAIHTGTILG--LENLSQSLDKNLTSEIPSOALENIPLELDSIGVNRHIT 237
 QY 137 VSKANSV-----FEDLSVTLROLR--NRLFOENSYSLSPL--NSLSRNNEDLLFSEL 187
 Db 238 ISR-NSLPDLNLSLEVNOIRLIPSDSETPPLSLTYLGNLTLSTIDASKMFLHIGL 296
 QY 188 QVL-----HDISSLSRKH-----HLAKDHPDLYSLAGLDEIGKRYGEDSE 231
 Db 297 KVLMSNNKIDITSIQNGKLSVOOFKLCITCIIAFOHAPSLIRLELDFPC----- 344
 QY 232 QFRDASKILVDAQKPADMYSLVGNNAVVELTVKVSFDTSLIKKRTTILEAKQAKNPAS 291

Db 264 AGVYHNDI-----KPGNVYEDRAS-----GE--PVYIDLGLHSRSGEPKGF--ESFKAPE 311

Qy 63 LAGNLEHRRRAVYMY-----KGVKALLPFGSVISPLENVP 103

Db 312 LGVNLASAKSEKSVFLVYVSTLLHCIESFEKNEPKRNGLEPITSEPAHMD---ENGYR 368

Qy 104 FSLDSVANSIHSLSFSEETPVYVLA-----PSEERYVWVGKANSVEDLSVTLRQLNR 157

Db 369 IHRGIA-GVETAYTRFTIDILGVASADSRPDSNEARLHFLSDGTIDEFSAKOI--LKDT 425

Qy 158 LFQENSVLSPLNSLRNNEVDLFLSELQVLIHOTSLSRHHKLAKDHPDL-----211

Db 426 LTGEMSPJST-----DVRRTTP-KKLREISDLRFLHLSAATKQIDMGCVLSD 472

Qy 212 YSLEIAGLDEIGRKGEDESEQFRDASKILVDALQKPADMYSLYGNAVVELTVKSPDT 271

Db 473 LDTMLVALDKAREGVDKDO-----LKSFN- 498

Qy 272 SLIRKRTILE-----AKQANKPASPYNLAKYKNEYSVY 306

Db 499 SLILKTYRVIEDYVKGREGDTKNSSTEVSPY---HRSNFMLSIV 539

RESULT 11

S30060

C:Species: Yersinia pseudotuberculosis

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999

C:Accession: S30060

R:Galiov, E.E.; Hakansson, S.; Forsberg, A.; Wolf-Matz, H.

Nature 361, 730-732, 1993

A:Title: A secreted protein kinase of Yersinia pseudotuberculosis is an indispensable virulence factor

A:Reference number: S30060; MUID:93180911

A:Accession: S30060

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-732 <DAL>

A:Cross-references: EMBL:X69439; NID:q49186; PIDN:CAA49215.1; PID:q49187

C:Keywords: serine/threonine-specific protein kinase

Query Match 5.9%; Score 103.5; DB 2; Length 732;

Best Local Similarity 21.2%; Pred. No. 10;

Matches 73; Conservative 51; Mismatches 103; Indels 117; Gaps 17;

Qy 12 AGVLGNFESILKSPGSVFRNGMPPIGERIPDYALSM-----GFSVKEDELSPWG 62

Db 264 AGVYHNDI-----KPGNVYEDRAS-----GE--PVYIDLGLHSRSGEPKGF--ESFKAPE 311

Qy 63 LAGNLEHRRRAVYMY-----KGVKALLPFGSVISPLENVP 103

Db 312 LGVNLASAKSEKSVFLVYVSTLLHCIESFEKNEPKRNGLEPITSEPAHMD---ENGYR 368

Qy 104 FSLDSVANSIHSLSFSEETPVYVLA-----PSEERYVWVGKANSVEDLSVTLRQLNR 157

Db 369 IHRGIA-GVETAYTRFTIDILGVASADSRPDSNEARLHFLSDGTIDEFSAKOI--LKDT 425

Qy 158 LFQENSVLSPLNSLRNNEVDLFLSELQVLIHOTSLSRHHKLAKDHPDL-----211

Db 426 LTGEMSPJST-----DVRRTTP-KKLREISDLRFLHLSAATKQIDMGCVLSD 472

Qy 212 YSLEIAGLDEIGRKGEDESEQFRDASKILVDALQKPADMYSLYGNAVVELTVKSPDT 271

Db 473 LDTMLVALDKAREGVDKDO-----LKSFN- 498

Qy 272 SLIRKRTILE-----AKQANKPASPYNLAKYKNEYSVY 306

Db 499 SLILKTYRVIEDYVKGREGDTKNSSTEVSPY---HRSNFMLSIV 539

RESULT 12

C84538

probable LRR receptor protein kinase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: C84538

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Unaiyam, L.; Tallon, euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: C84538

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-925 <STO>

A:Cross-references: GB:AEO02093; NID:q4544402; PIDN:AAD22312.1; CSPDB:GN00139

C:Genetics:

A:Gene: At2g16250

A:Map position: 2

Query Match 5.9%; Score 103.5; DB 2; Length 925;

Best Local Similarity 22.6%; Pred. No. 14;

Matches 86; Conservative 43; Mismatches 126; Indels 125; Gaps 16;

Qy 4 FVYLL-----ALVAGVLGNFES-----ILKSPGSVFRNGMPPIGERIPD-----44

Db 11 FVYLLLCVLVFPDCCVVGQTOSEKRLILNLRSLSGLRGDPWPKGDCVDMGICQE 70

Qy 45 -----VAALSMGFSV-----KEDLSWPGIANGNLFIRPRAT 75

Db 71 NCSIIIGINISGFRRTIRIGLNPFSVDPRLNLRISYFNASGLALP-TLPEMF-----123

Qy 76 VWWVWVGYNKALPFGSVISYPLENAVPSLDSVA-----NSIHSLSFSEETPVYV 126

Db 124 -----GVSLIALEVIDLSSCSVNGVYPTFLGNLTSLRTLNLSQNSLSLVSPIGOLLN 177

Qy 127 LAPSE-ERYVWVGKANSVEDLS--VTLRQLNR-----FQENSVL 165

Db 178 LSQDLSRNSFTGVLPQSFSSKLNLTLVDVSSVYLTGPIPGGLGALSKLHLNFSNSFS 237

Qy 166 SSLP-----LNSLSRNNNEVDLFLSELQVLIHOTSLSRHHKLAKDHPD 210

Db 238 SPIPELGLDVLNVLDFEDLSINLSGVSPOGLRLSKLQIMAGNLLSGTLPV-----D 291

Qy 211 LYSLEIAGLDEIGRKGEDESEQFRDAS-----KILVDALQKPADMYSLYGNAVVEL 263

Db 292 LFSAE-SQLOTLVLRNFGSGSLPDVCMSPKLRLIIDLAKNNFTGLPYSSYDSQDIAM 350

Qy 264 VTKSFSD-----TSLIRKTR 278

Db 351 VDSSNFTYGEELPIIRRR 370

RESULT 13

A71685

guanosine pentaphosphate phosphohydrolase (gppa) RP294 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000

C:Accession: A71685

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630; MUID:99039499

A:Accession: A71685

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-470 <NAD>

A:Cross-references: GB:AJ235271; GB:AJ235269; NID:q3868717; PIDN:CAA14755.1; PID:q3868

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: gppa; RP294

Query Match 5.8%; Score 103; DB 2; Length 470;

Best Local Similarity 21.7%; Pred. No. 5.8;
Matches 65; Conservative 56; Mismatches 123; Indels 56; Gaps 12;

```

QY 45 VVALSMGSPVDELSPGLAVGNLFHRRATVMMKCYNKL-ALPGSVISTPLENAV 103
DB 117 ISGIDAGCIYADLGGSLLELAHYN-----KKVGKIKSLPLGTQI----- 157
QY 104 FSLDSVANS-----IHSLFSEETPVYQLAPSEERYVMGKA-----NSVFEDSVTL 151
DB 158 -----IANSNSDVGLTKMLEEFGVA-----HYNVLTLGICALRLMSRIYAESINYL 207
QY 152 ROLNRRLPQENSYSLSPLNLSRNNEVDLLFSLQVYHDSLSLHKLAKDHPDL 211
DB 208 KULNH--PELRYVEFELYLEKLSQIDKLKLSYEOKAIVNNAVLT---KAMTKVFSPEK 262
QY 212 YSELTAGLDELGRYGEDEOFRDASKIL---VDALOKFADDMYSLYCGNAVELVTWK 267
DB 263 IISNVGLKE-GVRF--DGLPYHETEKDIYERVKRLVFNDRNICIKTEKYIEALQYLLIN 319
QY 268 SDDTSILKRTITLLEAKOKN--PASPYNLAYRYNEYSVFNWVIMIALAVITTS 325
DB 320 SDATLLITIELALMLAOKYKNIDKTLANFVSEFILSDIPFSHRQRLMGLALVTYTA 379

RESULT 14
J00894
P115 protein - Mycoplasma hyorhinis
C:Species: Mycoplasma hyorhinis
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 02-Feb-2001
C:Accession: J00894
R:Notaricola, S.M.; McIntosh, M.A.; Wise, K.S.
Gene 97, 77-85, 1991
A:Title: A Mycoplasma hyorhinis protein with sequence similarities to nucleotide-binding
A:Reference number: J00894; MUID:91138990
A:Accession: J00894
A:Molecule type: DNA
A:Residues: 1-979 <NOT>
A:Cross-references: GB:M34956
A:Note: The authors translated the codon AAA for residue 956 as Leu
C:Comment: This protein is located in the cytoplasm.
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: chromosome segregation protein SMC1
C:Keywords: nucleotide binding; P-loop
F;3239/Region: nucleotide-binding motif A (P-loop)

```

Query Match 5.8%; Score 103; DB 2; Length 979;
Best Local Similarity 19.6%; Pred. No. 17;
Matches 67; Conservative 63; Mismatches 126; Indels 86; Gaps 15;

```

QY 11 VAVYLGNEPSILKSPGVVFNQWPIRGRIPLDVAALSMGFSYKE-----D 57
DB 537 IAKLIDHKYIVSLBED-LFRPGGTITGSGSKLEKTSILNYDIKIKETNTLKFADDOHD 595
QY 58 LSWBGLAVGNLFHRRATVMM-----VMVGVNKLALPPGSVISTPLEN--A 101
DB 536 LKIQOQTYNEIEFTVSTIQOVKTEANSINSKUNILNELNKLANSEIFKQODES 655
QY 102 VPFSLDSVANSIHSLSFSEETPVYQLAPSEERYVM-----GKANSVFEDLSVTLQOLNR 157
DB 656 LNLSPDEKILNKEQISTLT---TELSSKKDRJLNLISEGKGETKKQELDAKRLRLNTQ 712
QY 158 LFOENSYSLSPLNLSRNNEVDLLFSLQVYHDSLSLHKLAKDHPDLVSLTA 217
DB 713 -----HSDSTIPEONRAKFL-----VEQNKRLSEHYKILTEASEQYSLD-- 752
QY 218 GLDIEGRYGEDE--EGRFDRASKILVDALOKFADDMYSLYGNA--VELVTVKS----- 268
DB 753 -LDIEGRYGEDE--EGRFDRASKILVDALOKFADDMYSLYGNA--VELVTVKS----- 268
QY 269 ---FDTSILKRTITLLEAKOKN--PASPYNLAYRYNEYSVFNWVIMIALAVITTS 325
DB 320 SDATLLITIELALMLAOKYKNIDKTLANFVSEFILSDIPFSHRQRLMGLALVTYTA 379

```

DB 811 ISDLDKIILINKTEIY-----NLV--NNERNWVF 837

```

RESULT 15
A39611
Probable GTP-binding protein - mouse
N:Alternate names: gbl10/Mov protein
C:Species: Mus musculus (house mouse)
C>Date: 13-Sep-1991 #sequence_revision 13-Sep-1991 #text_change 02-Feb-2001
C:Accession: A39611; I48358; S38628
R:Mooslechner, K.; Mueller, U.; Katis, U.; Hamann, L.; Harbers, K.
Mol. Cell. Biol. 11, 886-893, 1991
A:Title: Structure and expression of a gene encoding a putative GTP-binding protein 1
A:Reference number: A39611; MUID:91117255
A:Accession: A39611
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1004 <MOO>
A:Cross-references: EMBL:X52574; NID:953168; PIDN:CA436803.1; PID:953169
R:Hamann, L.; Bayer, K.U.; Jensen, K.; Harbers, K.
Mol. Cell. Biol. 14, 5786-5793, 1994
A:Title: Interaction of several related GC-box- and GT-box-binding proteins with the
A:Reference number: I48358; MUID:94344088
A:Accession: I48358
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <RES>
A:Cross-references: EMBL:X75819; NID:9415391; PIDN:CA453453.1; PID:9433685
C:Genetics:
A:Gene: gbl10
C:Keywords: nucleotide binding; P-loop
F;525-532/Region: nucleotide-binding motif A (P-loop)

```

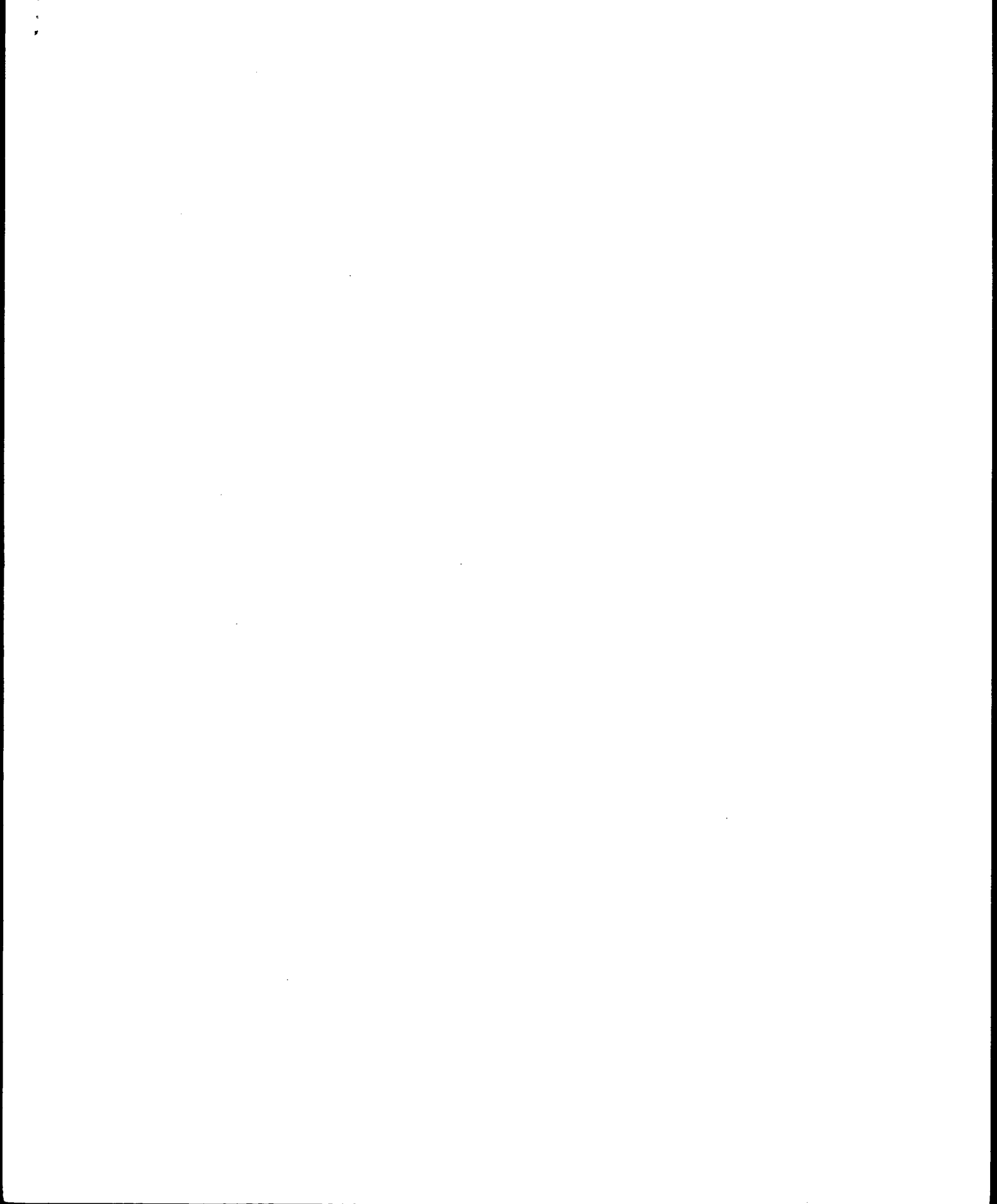
Query Match 5.8%; Score 102.5; DB 2; Length 1004;
Best Local Similarity 27.9%; Pred. No. 19;
Matches 56; Conservative 19; Mismatches 67; Indels 59; Gaps 9;

```

QY 149 VTLQOLNRLFOENSVLSS-----LPLNSLRNNEVD--LLFSLQVYHDS 195
DB 299 ILKQOLLPTLLQPSIFTAPEKVAEIKQLETTLSRNVEYKRLHLEELQEMHDP-- 356
QY 136 LLSHKLAK-----DHSPLYSLELAGDELGRYGEDEOFRDASKILVDALOKF 247
DB 357 ---RHVDLSVPMTWDPVDONPRLTLLEVPVGAESRPSVLKGDHLF---ALLSEETQO- 408
QY 248 ADDWYSLYGNAVVELVTWK--SPDTSILKRTITLLEAKOKN--PASPYNLAYRYNEYS-- 304
DB 409 -DDVYTYKGVYKVELDVKLSFSTLSLRVD-----GLTFKVNFTFRQ 453
QY 305 -----VVENNVIMIMT 315
DB 454 PLRVQRALELTGRWVLPML 474

```

Search completed: July 3, 2001, 16:19:18
Job time: 155 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:17:23 : Search time 12.1 Seconds

(Without alignments)
990,861 Million cell updates/sec

Title: US-09-284-320-6

Sequence: 1 MAVFVVLALVAGVLGNFES.....MDPGVDIIYRMNOKIRMD 350

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swisprot_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	513	29.1	1 VATN_HUMAN	O75787 homo sapien
2	122	6.9	30 VATN_BOVIN	P81134 bos taurus
3	105	6.0	1 DNA_SPIC1	P34028 spiroplasma
4	103.5	5.9	502 Y190_HELPY	P56117 helicobacte
5	103.5	5.9	732 YPKA_YERPS	O05608 yersinia ps
6	103	979	1 P15_MYCHR	P41508 mycoplasma
7	102.5	5.8	1004 MVI0_MOUSE	P23249 mus musculu
8	100.5	5.7	1358 S1R4_YEAST	P11978 saccharomyc
9	100.5	5.7	2839 1 NEFL_HUMAN	P21359 homo sapien
10	99	5.6	483 1 SYD_MYCA	O48979 mycoplasma
11	98.5	5.6	1935 1 MYSB_HUMAN	P12883 homo sapien
12	97.5	5.5	884 1 SECA_PORPU	P51381 porphyra pu
13	97.5	5.5	2841 1 NEFL_MOUSE	O04690 mus musculu
14	97	5.5	715 1 ORC3_MOUSE	O91K30 mus musculu
15	96.5	5.5	805 1 SUSI_TULGE	O41608 tulipa gesn
16	96.5	5.5	1934 1 MYSB_MESAU	P13540 mesocricetu
17	96.5	5.5	1938 1 MYSB_RAT	P02563 rattus norv
18	95.5	5.4	736 1 MYSB_RABIT	P04461 onycholagus
19	94.5	5.4	1935 1 MYSB_PIG	P79293 sus scrofa
20	94.5	5.4	1935 1 MYSB_RAT	P02564 rattus norv
21	93.5	5.3	1939 1 MYSA_HUMAN	P13533 homo sapien
22	93	640	1 DKS_SYNY3	P73037 synechocyst
23	92.5	5.2	502 Y190_HELPY	O9amp2 helicobacte
24	91.5	5.2	896 1 BOSS_DROME	P22815 drosophila
25	91.5	5.2	4092 1 DYHC_YEAST	P36022 saccharomyc
26	91	5.2	262 1 CUT8_SCHPO	P38937 schizosach
27	91	5.2	427 1 FOLD_YEAST	Q12676 saccharomyc
28	90.5	5.1	465 1 MYSA_RABIT	P04460 onycholagus
29	90.5	5.1	472 1 ATPB_RHOCA	P72247 rhodobacter
30	90.5	5.1	1102 1 MYSC_CHICK	P29616 gallus gall
31	90.5	5.1	1938 1 MYSA_MOUSE	O02566 mus musculu
32	90	5.1	507 1 RECN_CAMEL	O9pums camylobact
33	90	5.1	626 1 CEIB_ECOLI	P04479 escherichia

34	90	5.1	1032 1 KINN_HUMAN	Q12840 homo sapien
35	90	5.1	1085 1 CUT7_SCHPO	P24339 schizosach
36	89.5	5.1	642 1 DEAD_KLEPN	P31906 klebsiella
37	89.5	5.1	968 1 Y682_METJA	O58095 methanococc
38	89.5	5.1	3106 1 LMA2_MOUSE	O60675 mus musculu
39	89	5.0	500 1 TACY_CLOPE	P19995 clostridium
40	89	5.0	904 1 PMS1_YEAST	P14242 saccharomyc
41	88.5	5.0	1663 1 CO3_HUMAN	P01024 homo sapien
42	88	5.0	442 1 THDE_MYCE	P47254 mycoplasma
43	88	5.0	618 1 VE1_HPV63	O07847 human papil
44	88	5.0	1177 1 MFD_BACSU	P37474 bacillus su
45	87.5	5.0	505 1 HAS1_YEAST	O03532 saccharomyc

ALIGNMENTS

RESULT 1	VATN_HUMAN	STANDARD:	PRT:	100 AA.
AC	O75787:			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	VACUOLAR ATP SYNTHASE MEMBRANE SECTOR ASSOCIATED PROTEIN M8-9			
DE	(V-ATPASE M8.9 SUBUNIT).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OX	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
RN	NCBI_Taxid=9606;			
RP	SEQUENCE FROM N.A.			
RA	Luckaig J., Schagger H., Boyd A., Apps D.K.			
RL	Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.			
CC	- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY			
CC	OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.			
CC	- SUBUNIT: COMPOSED OF AT LEAST 10 SUBUNITS.			
CC	- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: Y17975: CAA76984.1; -			
KW	Hydrolase; Hydrogen Ion transport; Transmembrane.			
FT	TRANSMEM 53 73 POTENTIAL.			
SQ	SEQUENCE 100 AA: 11575 MW: A0705DB7665F6DC CRC64:			

Query Match 29.1%: Score 513; DB 1; Length 100;
Best Local Similarity 100.0%: Pred. No. 1.6e-31;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	251 MYSLGNGAVELVYVYKSFDSILRKRTITLAKOAKNPASPYMLAKYKFEYVFMV 310	
DB	1 MYSLGNGAVELVYVYKSFDSILRKRTITLAKOAKNPASPYMLAKYKFEYVFMV 60	
OY	311 LMIMIALALAVIITSYNIWMNDPGYDSIIYRMNOKIRMD 350	
DB	61 LMIMIALALAVIITSYNIWMNDPGYDSIIYRMNOKIRMD 100	

RESULT 2	VATN_BOVIN	STANDARD:	PRT:	30 AA.
ID	VATN_BOVIN			
AC	P81134:			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			

DE VACUOLAR ATP SYNTHASE MEMBRANE SECTOR ASSOCIATED PROTEIN M8-9
 (V-ATPASE M8.9 SUBUNIT) (FRAGMENT).
 OS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovinae; Bovinae; Bos.
 CC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Adrenal medulla;
 RX MEDLINE=98225166; PubMed=9556572;
 RA Ludwig J., Kerscher S., Brandt U., Pfeiffer K., Gellera F., Apps D.K.,
 RT Schagger H.,
 RT "Identification and characterization of a novel 9.2-kDa membrane
 RT sector-associated protein of vacuolar proton-ATPase from chromaffin
 RT granules.";
 RL J. Biol. Chem. 273:10939-10947(1998).
 CC -I- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
 CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
 CC -I- SUBUNIT: COMPOSED OF AT LEAST 10 SUBUNITS.
 CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC Hydrolyase; Hydrogen ion transport; Transmembrane.
 FT NON_TER 1 1
 FT TRANSMEM 22 >30 POTENTIAL.
 FT NON_TER 30 30
 SQ SEQUENCE 30 AA; 3585 MW; EA9FB26DCA7D665 CRC64;

Query Match 6.9%; Score 122; DB 1; Length 30;
 Best Local Similarity 73.3%; Pred. No. 0.0018;
 Matches 22; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 282 EAKQAKNPASPVNLAAYKYFYEVVFNMYL 311
 DB 1 ETKQVKDPSTYVLAAYKYFYEVVFNMYL 30

RESULT 3
 DNA_SPTCI STANDARD; PRT; 450 AA.
 ID P34028;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNA.
 GN DNA.
 OS Spiroplasma citri.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Spiroplasmataceae; Spiroplasma.
 CC NCBI_TaxID=2133;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 27556 / R8A2;
 RX MEDLINE=94290261; PubMed=7764984;
 RA ye F., Renaudin J., Bove J.M., Laigret F.,
 RT "Cloning and sequencing of the replication origin (oric) of the
 RT Spiroplasma citri chromosome and construction of autonomously
 RT replicating artificial plasmids.";
 RT Curr. Microbiol. 29:23-29(1994).
 RN [2]
 RP SEQUENCE OF 147-411 FROM N.A.
 RC STRAIN=ATCC 27556 / R8A2;
 RX MEDLINE=94109679; PubMed=8282191;
 RA Suzuki K., Miyata M., Fukumura T.,
 RT "Comparison of the conserved region in the dnaA gene from three
 RT mollusc species.";
 RL FEBS Microbiol. Lett. 114:229-233(1993).
 CC -I- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION
 CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;
 CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS
 CC (DNA BOX): 5'-TTATC(C/A)A(C/A)A-3'. DNA BINDS TO ATP AND TO
 CC ACIDIC PHOSPHOLIPIDS.
 CC -I- SIMILARITY: BELONGS TO THE DNAA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 DR EMBL: Z19108; CAA79521.1; -;
 DR EMBL: D14985; BAA03630.1; -;
 DR PIR: S35732; S35732.
 DR InterPro: IPR001957; -;
 DR Pfam: PF00308; bac_dnaA; 1.
 DR PRINTS: PR00051; DNAA.
 DR PROSITE: PS01008; DNAA; 1.
 KW DNA replication; DNA-binding; ATP-binding.
 FT NP_BIND 146 153
 SQ SEQUENCE 450 AA; 51537 MW; C401EF9B157FB961 CRC64;

Query Match 6.0%; Score 105; DB 1; Length 450;
 Best Local Similarity 20.4%; Pred. No. 1.6;
 Matches 67; Conservative 41; Mismatches 115; Indels 106; Gaps 13;

QY 81 KGVNKLALPPGVSIVS-----YPLENAVPSLSVANSIHISFSEEPVY 124
 DB 83 KNINKQA-----SVISKIDLTENNNAIYENTFENFVR-----GDSNHEAKQALAVA 132
 QY 125 LQ LAPSEERYVWYGKANSVFEDLSVTLROLNRRLFOENS-----VLSSLPINLSRN 176
 DB 133 LDLGKKMNPFLFYGGSGLGKTHLHAENKRYEIKYTNKRYKAKDEFKGIAMDINOG 192
 QY 177 NEV-----DLFLSELQV-----HDISLSLRKH--LAKHSPDLY 212
 DB 193 HEIIEAKTSYDIOCLHDIQLAKRKKTNELFFHFNSTYIEKKQIVITSKYD-- 250
 QY 213 SLEIAGD-----ELGKRYGDSQFPDASKIIVDALQ-----KF 247
 DB 251 --DLGGFEARIISRFSYGLSIGDSDPFETALKLEOKLKHNNGLFSESLFIFALNF 308
 QY 248 ADDWVSLYGG-----NAVELVTV--KSPDTSIRKTRTLEKQAKNPASP 292
 DB 309 NSVDKRLIEGAKIRLLEFLAVMKKRPNELITLADVEAKRKNAPIONNKKITPKKIQIVADS 368
 QY 293 YNLAYKRYEYSVFENWYIMIALAV 321
 DB 369 YNIIKAMMSKRSVSN-----VMOAROLAM 393

RESULT 4
 Y190_HELPY STANDARD; PRT; 502 AA.
 ID Y190_HELPY
 AC P56117;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL PROTEIN HP0190.
 GN HP0190.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 CC Helicobacter.
 CC NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,

RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venner J.C.;
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*
 RT *pylori*.";
 RL Nature 388:539-547(1997).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN
 CC SYNTHASE SUBFAMILY. STRONG, TO E.COLI YMDC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000539; AAD07257.1; -;
 DR TIGR: HP0190; -;
 DR InterPro: IPR001736; -;
 DR Pfam: PF00614; PLDC; 2;
 KW Hypothetical protein; Transferase; Transmembrane.
 FT TRANSMEM 1
 FT POTENTIAL 21
 SQ SEQUENCE 502 AA; 58287 MW; C80911F2B59C714 CRC64;

Query Match 5.9%; Score 103.5; DB 1; Length 502;
 Best Local Similarity 22.9%; Pred. No. 2.5;
 Matches 72; Conservative 35; Mismatches 115; Indels 93; Gaps 15;

QY 93 VTSPLRNA-----VPSLDSVANSIHSLSFSEPPVVIOLAPSEEV-----YWG 138
 DB 95 VIAELNANNGVKKYRLILD--NGIDSPFS-----IMLNHKNIEKIFNPYIRN 147
 QY 139 KANSVE--DISVTLROLRRRLFOENSVLSLPLNSLRN--NEVDLFLSELQV-- 190
 DB 148 KGLYFEMLADYERIKKRNHKLFTVDNFVILIGGNIGDVFENDDTNFL-DLDALEF 206
 QY 191 -----HDS-SLSRKHAKHSDPDLSTLGLADGCKRGGSE 231
 DB 207 GGVAKKAKSEFERYWFRHSIPVSLTNRKLNK-----AKIALNHEKIPISAEKN 260
 QY 233 QFRDASILVDALQFADDMYSLYGNAVVELVYKSFDTSLIRKRTTILEAKQKNPAS 291
 DB 261 QPEKKVNDIFDFOKY--QYPIYGNAIFLADSPKIDTFLYPIKIAFE-KALKNAKD 316
 QY 292 PYNLATYV-----NFEYSVVENMYLMIAMIALAVIITSYNNMMPG 334
 DB 317 SVFLASSYFIPGKKMMKIFKNOISKIETLNLIN-----SLSDTAIVYGAME---- 365
 QY 335 YDSIYRMTNOKIRM 349
 DB 366 -----RYRNQLVRM 374

RESULT 5
 YPKA_TERPS STANDARD: PRT; 732 AA.
 ID YPKA_TERPS
 AC 005608;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PROTEIN KINASE YPKA PRECURSOR (EC 2.7.1.1).
 GN YPKA.
 OS *Yersinia pseudotuberculosis*.
 OG *Yersinia* pIBL.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Yersinia*.
 OX NCBI_TaxID=633;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YPIII;

RX MEDLINE=93180911; PubMed=8441468;
 RA Galyov E.E., Hakansson S., Forsberg A., Wolf-Watz H.;
 RT "A secreted protein kinase of *Yersinia pseudotuberculosis* is an
 RT indispensable virulence determinant";
 RL Nature 361:730-732(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YPIII;
 RX MEDLINE=94321324; PubMed=8045884;
 RA Galyov E.E., Hakansson S., Wolf-Watz H.;
 RT "Characterization of the operon encoding the YpkA Ser/Thr protein
 RT kinase and the YopJ protein of *Yersinia pseudotuberculosis*.";
 RL J. Bacteriol. 176:4543-4548(1994).
 CC -1- FUNCTION: ACTS AS A VIRULENCE DETERMINANT.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X69439; CAA49215.1; -;
 DR EMBL: L33833; AAA68487.1; -;
 DR PIR: S3060; S3060.
 DR InterPro: IPR000719; -;
 DR InterPro: IPR002290; -;
 DR PROSITE: PS00107; PROTEIN KINASE ATP; FALSE_NEG.
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 KW Virulence; Plasmid; Transferase; Serine/threonine-protein kinase;
 KW ATP-binding; Signal.
 FT SIGNAL 1
 FT CHAIN 2
 FT DOMAIN 136 408
 FT ACT_SITE 270 270
 FT BY_SIMILARITY
 SQ SEQUENCE 732 AA; 81707 MW; 7C09BDD4F9C9CBDF CRC64;

Query Match 5.9%; Score 103.5; DB 1; Length 732;
 Best Local Similarity 21.2%; Pred. No. 4.2;
 Matches 73; Conservative 51; Mismatches 103; Indels 117; Gaps 17;

QY 12 AGVAGNEFSILKSPGSVFRNGNPIGEPRIQVAALSM-----GFSYKEDLSWPG 62
 DB 264 AGVYHNDI--KPGVVFDDAS--GE--PVYIDLGHSRSGEOPKGT--ESFKAPE 311
 QY 63 LAVGNLFRPRATVWV-----KGVNKLALPGSVISTPLENAV 103
 DB 312 LGVNLGASERKSDVFLVSTLLHCIEGFEKNEPKRNOGLRFLITSEPAHYMD--ENQYP 368
 QY 104 FSLDSVANSIHSLSFSEPPVVIOLA-----PSEERYVMGKANSVEDLSTLRQLNR 157
 DB 369 IHRPGIA-GVETAYTRFTIDILGVSADSRDSEARLHEPLSDTIDESAQOI-LKDT 425
 QY 158 LFQENSVLSLPLNSLRNNEVDLFLSELQVLDITSSLSRKHAKHSDPDL----- 211
 DB 426 LTGEKSPLSLT-----DVRKITP-KKLRLSLDLTRTHLSAATKOLDMGVLSD 472
 QY 212 YSLTAGIDELGKRGVDESEDFRASKIIVDALQKADDMYSLYGNAVVELVYKSFDT 271
 DB 473 LDTMIVALDKAREGGVKKDQ-----LAKSPN- 498
 QY 272 SLIRKRTILE-----AKQAKNDAPYNLAKYNEYSV 306
 DB 499 SLIKTYRVIEDYVKGREGDTKNSSTEVSPY--HRSNFMLSIV 539

RESULT 6
 P115_MYCHR


```

RESULT 8
SIR4_YEAST STANDARD; PRT; 1358 AA.
ID SIR4_YEAST
AC P11978;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE REGULATORY PROTEIN SIR4 (SILENT INFORMATION REGULATOR 4).
GN SIR4 OR STE9 OR ASD1 OR UTH2 OR YDR227W OR YD9934.12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88142836; PubMed=3325825;
RA Marshall M., Mahoney D., Rose A., Hicks J.B., Broach J.R.;
RT "Functional domains of SIR4, a gene required for position effect
RT regulation in Saccharomyces cerevisiae";
RL Mol. Cell. Biol. 7:4441-4452(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95192063; PubMed=7885847;
RA Davies C.J., Hutchinson C.A. III;
RT "Insertion site specificity of the transposon Tn3.";
RL Nucleic Acids Res. 23:507-514(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A.;
RT Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THE PROTEIN SIR1 THROUGH SIR4 CONTROL THE EXPRESSION OF
CC MATING TYPE GENES RESIDENT AT LOCI OF EITHER END OF CHROMOSOME
CC III. SIR3 AND SIR4 ASSOCIATE WITH THE C-TERMINUS OF RAP1 TO FORM A
CC DNA-BINDING COMPLEX THAT INITIATES THE REPRESSION AT THE HM LOCI
CC AND TELOMERES.
CC -1- SUBUNIT: INTERACTS WITH RAP1 C-TERMINUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: M37249; AAA20881.1; -
DR EMBL: U13239; AAC33144.1; -
DR EMBL: Z48612; CAA88507.1; -
DR PIR: A29360; A29360.
DR SGD: S0002635; SIR4.
KW transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW Coiled coil.
FT DOMAIN 1277 1347 COILED COIL (POTENTIAL).
FT VARIANT 994 994 P -> L.
SQ SEQUENCE 1358 AA; 152061 MW; 9C698765964F094E CAC64;

Query Match 5.7%; Score 100.5; DB 1; Length 1358;
Best Local Similarity 24.1%; Pred. No. 17;
Matches 70; Conservative 43; Mismatches 92; Indels 85; Gaps 17;

OY 38 PGERIPVAVALS-----MGESYKEDLSWPLANGNLPHPRATVWVWVKGVNKIA 87
DB 23 PNDKTFPEEKSEKNEVTKPIPLFTFAKSKVYSRPAIHTSPROPSD--VKFTSHKOLQ 79
OY 88 LPPGSVISTYLENAVPS-LOSIVANS-IHSLFSEETPVVLOLAPSEERYVWGKANSVEE 145
DB 80 QPKSSPLKKNVNSFPSPHSLKISKLISLRSKT-----SAGRI-----ESNSNSH 127
OY 146 DLSVTLQLRLNRLF-----QENSVLSS-----LPLNSLRNNEVDLLFSELOVLDHS 194
DB 128 DASRSLASFQDTAFSRHADQQTSTFNKSPVRIIVTISISQINN-----FLS-----GVK 177

OY 195 SLLSRHKHLAKDHSPLYSLELAGDEIGKRRYGEDESEQFRDASKILLVDALQFADMYSL 254
DB 178 SLISSEK--IRDYSKELIGINLA-----NEO-----PVLEKPLKKGSAADI--- 215
OY 255 YGSAVAVELVWVSFDTSLIKRTTILEAKOKANPASPYNLAKYNEFS 304
DB 216 --GASVSLTKKDS-----IRKD-TYFEKKER-----LNTGKNFANS 250

RESULT 9
NFI_HUMAN STANDARD; PRT; 2839 AA.
ID NFI_HUMAN
AC P21359;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).
GN NFI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-1370 AND 1392-2839 FROM N.A.
RX MEDLINE=92147138; PubMed=1783401;
RA Marchuk D.A., Saulino A., Tavakkol R., Swaroop M., Wallace M.R.,
RA Andersen L.B., Mitchell A.L., Gutmann D.H., Boguski M., Collins F.S.;
RT "cDNA cloning of the type 1 neurofibromatosis gene: complete sequence
RT of the NFI gene product.";
RL Genomics 11:931-940(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93090270; PubMed=1457041;
RA Bernards A., Haase V.H., Murthy A.E., Menon A., Hannigan G.E.,
RA Gusella J.F.;
RT "Complete human NFI cDNA sequence: two alternatively spliced mRNAs
RT and absence of expression in a neuroblastoma line.";
RL DNA Cell Biol. 11:727-734(1992).
RN [3]
RP SEQUENCE OF 335-1370 AND 1392-2839 FROM N.A.
RX MEDLINE=90335969; PubMed=2116237;
RA Xu G., O'Connell P., Viskochil D., Cawthon R., Robertson M.,
RA Culver M., Dunn D., Stevens J., Gesteland R., White R.;
RT "The neurofibromatosis type 1 gene encodes a protein related to GAP.";
RL Cell 62:599-608(1990).
RN [4]
RP SEQUENCE OF 1096-1370 AND 1372-1590 FROM N.A.
RX MEDLINE=91029515; PubMed=2121370;
RA Martin G.A., Viskochil D., Bollag G., McCabe P.C., Crosier W.J.,
RA Haubruck H., Conroy L., Clark R., O'Connell P., Cawthon R.M.,
RA Innis M., McCormick F.;
RT "The GAP-related domain of the neurofibromatosis type 1 gene product
RT interacts with ras p21.";
RL Cell 63:843-849(1990).
RN [5]
RP SEQUENCE OF 1606-2709 FROM N.A. AND VARIANT PRO-1953.
RX MEDLINE=90304909; PubMed=2114220;
RA Cawthon R.M., Weiss R., Xu G., Viskochil D., Culver M., Stevens J.,
RA Robertson M., Dunn D., Gesteland R., O'Connell P., White R.;
RT "A major segment of the neurofibromatosis type 1 gene: cDNA sequence,
RT genomic structure, and point mutations.";
RL Cell 62:193-201(1990).
RN [6]
RP SEQUENCE OF 2230-2839 FROM N.A.
RX MEDLINE=90319792; PubMed=2134734;
RA Wallace M.R., Marchuk D.A., Andersen L.B., Letcher R., Odeh H.M.,
RA Saulino A.M., Fountaine J.W., Breton A., Nicholson J., Mitchell A.L.,
RA Brownstein B.H., Collins F.S.;
RT "Type 1 neurofibromatosis gene: identification of a large transcript
RT disrupted in three NFI patients.";
RL Science 249:181-186(1990).
RN [7]

```

RP ERRATUM. PubMed=2125369.
RX MEDLINE-91102559; PubMed=2125369.
RA Wallace M.R., Marchuk D.A., Andersen L.B., Collins F.S.;
RL Science 250:1749-1749(1990).
RN [8]
RT
RT SEQUENCE OF 1168-1566 FROM N.A.
RX MEDLINE-92019823; PubMed=1923522;
RA Nishl T., Lee P.S., Oka K., Levin V.A., Tanase S., Morino Y.,
RT Saya H.;
RT "Differential expression of two types of the neurofibromatosis type 1
RT (NF1) gene transcripts related to neuronal differentiation.";
RL Oncogene 6:1555-1559(1991).
RN [9]
RP SEQUENCE OF 1371-1391 FROM N.A.
RX MEDLINE-93109335; PubMed=8417346;
RA Andersen L.B., Ballaster R., Marchuk D.A., Chang E., Gutmann D.H.,
RT Saulino A.M., Camonis J., Wigler M., Collins F.S.;
RT "A conserved alternative splice in the von Recklinghausen
RT neurofibromatosis (NF1) gene produces two neurofibrin isoforms,
RT both of which have G-protein-activating protein activity.";
RL Mol. Cell. Biol. 13:487-495(1993).
RN [10]
RP FUNCTION.
RX MEDLINE-91029516; PubMed=2121371;
RA Ballaster R., Marchuk D., Boguski M.S., Saulino A., Letcher R.,
RT Wigler M., Collins F.S.;
RT "The NF1 locus encodes a protein functionally related to mammalian
RL GAP and yeast IRA proteins.";
RL Cell 63:851-859(1990).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE-95072625; PubMed=7981724;
RA Upadhyaya M., Shaw D.J., Harper P.S.;
RT "Molecular basis of neurofibromatosis type 1 (NF1)
RT and polymorphisms in the NF1 gene.";
RL Hum. Mutat. 4:83-101(1994).
RN [12]
RP REVIEW ON VARIANTS.
RX MEDLINE-96422425; PubMed=8825042;
RA Hong Shen M., Harper P.S., Upadhyaya M.;
RT "Molecular genetics of neurofibromatosis type 1 (NF1).";
RL J. Med. Genet. 33:2-17(1996).
RN [13]
RP VARIANT GU2-1444.
RX MEDLINE-92233464; PubMed=1568247;
RA Li Y., Bolivar G., Clark R., Stevens J., Conroy L., Puls D., Ward K.,
RT Friedman E., Samowitz W., Robertson M., Bradley P., McCormick F.,
RA White R., Cawthon R.;
RT "Somatic mutations in the neurofibromatosis 1 gene in human tumors.";
RL Cell 69:275-281(1992).
RN [14]
RP VARIANTS MET-2164 AND ASN-2192.
RX MEDLINE-93256316; PubMed=1302608;
RA Upadhyaya M., Shen M., Cherryson A., Farnham J., Maynard J.,
RT Huson S.M., Harper P.S.;
RT Analysis of mutations at the neurofibromatosis 1 (NF1) locus.";
RN Hum. Mol. Genet. 1:735-740(1992).
RN [15]
RP VARIANT HIS-1721-LEU-1733 DPPL.
RX MEDLINE-93304433; PubMed=8317503;
RA Tassabehji M., Strachan T., Sharland M., Colley A., Donnai D.,
RT Harris R., Thakker N.;
RT "Tandem duplication within a neurofibromatosis type 1 (NF1) gene exon
RT in a family with features of Watson syndrome and Noonan syndrome.";
RL Am. J. Hum. Genet. 53:90-95(1993).
RN [16]
RP VARIANT MET-991 DEL.
RX MEDLINE-94108439; PubMed=7904209;
RA Shen M.H., Harper P.S., Upadhyaya M.;
RT "Neurofibromatosis type 1 (NF1): the search for mutations by PCR-
RT heteroduplex analysis on Hydrilink gels.";
RL Hum. Mol. Genet. 2:1861-1864(1993).
RN [17]

RP VARIANT NF1 ASN-2387 --PHE-2388 DEL.
 RP MEDLINE-94362704; PubMed-8081387;
 RA Abernathy C.R., Colman S.D., Kousseff B.G., Wallace M.R.;
 RT "Two NF1 mutations: frameshift in the GAP-related domain, and loss of
 RT two codons toward the 3' end of the gene.";
 RL Hum. Mutat. 3:347-352(1994).
 RN (18)
 RP VARIANT NF1 ALA-2631.
 RP MEDLINE-96091873; PubMed-8544190;
 RA Upadhyaya M., Maynard J., Osborn M., Huson S.M., Ponder M.,
 RA Ponder B.A.J., Harper P.S.;
 RT "Characterisation of germline mutations in the neurofibromatosis type
 RT 1 (NF1) gene.";
 RL J. Med. Genet. 32:706-710(1995).
 RN (19)
 RP VARIANT NF1 ARG-629.
 RP MEDLINE-96431167; PubMed-8834249;
 RA Gasparini P., D'Aguma L., de Cillis G.P., Balestrazzi P.,
 RA Mingarelli R., Zelante L.;
 RT "Scanning the first part of the neurofibromatosis type 1 gene by RNA-
 RT SSC: identification of three novel mutations and of two new
 RT polymorphisms.";
 RL Hum. Genet. 97:492-495(1996).
 RN (20)
 RP VARIANT LS ARG-1035.
 RP MEDLINE-96400960; PubMed-8807336;
 RA Wu R., Legius E., Robberecht W., Dumoulin M., Cassiman J.-J.,
 RA Fryns J.-P.;
 RT "Neurofibromatosis type I gene mutation in a patient with features of
 RT LEOPARD syndrome.";
 RL Hum. Mutat. 8:51-56(1996).
 RN (21)
 RP VARIANTS NF1 ARG-844 AND PRO-898.
 RP MEDLINE-97295087; PubMed-9150739;
 RA Maynard J., Krawczak M., Upadhyaya M.;
 RT "Characterization and significance of nine novel mutations in exon 16
 RT of the neurofibromatosis type I (NF1) gene.";
 RL Hum. Genet. 99:674-676(1997).
 RN (22)
 RP VARIANT NF1 ARG-1952.
 RP MEDLINE-97255969; PubMed-9101300;
 RA Hudson J., Wu C.L., Tassabehji M., Summers E.M., Simon S., Super M.,
 RA Donnai D., Thakker N.;
 RT "Novel and recurrent mutations in the neurofibromatosis type I (NF1)
 RT gene.";
 RL Hum. Mutat. 9:366-367(1997).
 RN (23)
 RP VARIANT NF1 TRP-1611.
 RP MEDLINE-97442280; PubMed-9298829;
 RA Upadhyaya M., Maynard J., Osborn M., Harper P.S.;
 RT "Six novel mutations in the neurofibromatosis type I (NF1) gene.";
 RL Hum. Mutat. 10:248-250(1997).
 RN (24)
 RP VARIANT NF1 PRO-1446.
 RP MEDLINE-9925555; PubMed-10220149;
 RA Peters H., Hess D., Fahsold R., Schuelke M.;
 RT "A novel mutation G1429P in the GAP-region of the NF1 gene detected by
 RT temperature gradient gel electrophoresis (TGGE).";
 RL Hum. Mutat. 13:337-337(1999).
 CC CC
 CC -I- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NF1 SHOWS GREATER
 CC AFFINITY FOR RAS GTP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE
 CC A REGULATOR OF RAS ACTIVITY.
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS: I AND II (SHOWN HERE); ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -I- DISEASE: THIS PROTEIN IS ASSOCIATED WITH TYPE 1 NEUROFIBROMATOSIS
 CC (NF1) (ALSO CALLED VON RECKLINGHAUSEN SYNDROME), THE MOST FREQUENT

```

Db 2021 GGLSIAAEVMDPTVALASGNVSKYIGRMKIIDTKLSPPTLEOHLMMDIAI 2080
QY 66 GNLFRRPRAVWVWVGNKLAIPG-----SVISYPLENAVPESSIDS-----VANSIHS 115
Db 2081 -----LARYMLMLSFNNSIDVAHLPYLFIHVTF-LVATGPIRSTASTHOLVINIHS 2132
QY 116 L-----PSEEPVYLOLAPSE---ERYVMWGNKNSVEDELSTLR-QLRRRLQGENSV 164
Db 2133 LCTGSLHSEETKOYLRSLTEFLPKFYLFGISKYSKANAIAFRSSYKDRSPGSGY 2192
QY 165 -LSSPLNLSIRNNEVDLLFSELQVLDHDISSL-LSRHKHLAK---DHSPLDIYLELA 217
Db 2193 EREFEALTSLFETEA-LLEIIME-ACMRDIPCKMLDQETELAQRFAPQYNSLQPRALV 2250
QY 218 GLDIDIGR--YGEDESEOPRDSKTLVDALOKRADMS-----253
Db 2251 VFGCISKRSYHQIKQIIRILSKALESCUK--GPDITSOVLLEATVIALTKLOPLMKD 2308
QY 234 -----LYGNAVVEL--VTKSEFDTSLIRKRTILEAKQANKPASP-----292
Db 2309 SPLKALFWAAVAVQLDEVNLYSAGTALLEONLHTLDSLIRFNKSPREVMAIRNPLE 2368
QY 293 -----YNLAYKYNEYSVPEVNMV 310
Db 2369 WKKQMDHPVGLNFSNPNFALVGHLL 2395

RESULT 10
SYD_MYCCA
ID SYD_MYCCA STANDARD; PRT; 483 AA.
AC Q48979;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ASPRTYL-TRNA SYNTHETASE (EC 6.1.1.12) (ASPARTATE-TRNA LIGASE)
DE (ASPRS) (FRAGMENT).
GN ASPRS.
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Entomoplasmataceae.
OX NCBI_TaxID=2095;
RN [1]
RP SEQUENCE FROM N.A. / KID;
RC STRAIN=ATCC 27343 / KID;
RA MEDLINE=96059641; PubMed=7476192;
RA Bork P., Ouzounis C., Casari G., Schneider R., Sander C.,
RA Dolan M., Gilbert W., Gillevet P.M.;
RA "Exploring the Mycoplasma capricolum genome: a minimal cell reveals
RT its physiology.";
RL Mol. Microbiol. 16:955-967(1995).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + TRNA(ASP) = AMP +
CC PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, Z33048; CAA83720.1; -.
DR InterPro: IPR002106; -.
DR InterPro: IPR002309; -.
DR Pfam: PF00152; tRNA-synt_2; 1.
DR PROSITE: PS00179; AA-TRNA_LIGASE_II_1; 1.
DR PROSITE: PS00339; AA-TRNA_LIGASE_II_2; PARTIAL.
KM Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT NON_TER 1
SO SEQUENCE 483 AA; 56298 MW; A03F5B0CEA31087E CRC64;

```

```

Query Match 5.6%; Score 99; DB 1; Length 483;
Best Local Similarity 25.8%; Pred. No. 5;
Matches 62; Conservative 38; Mismatches 84; Indels 56; Gaps 14;

QY 76 VMWVWVGNKLAIPGVSISYPLENAVPESSIDVANSIHSIFSEEPVYLOLAPSEERY 135
Db 31 IYVKKSVNK-----ELITGEIVYKDL--VINK-----SELTPYLE-----68
QY 136 WGNKANSVEDELSTLR--QLRRRLQGENSVLSSPLNLSIRNNEVDLLFSELQVLDH- 192
Db 69 ---MDVAVNEDRLRYLRYDLDRQVQNNULIIRA-KINHIIIRVLDLNLFL-EVETPFYA 123
QY 193 -----ISSLSRHKHLAKDHSPLDYS--LELAGDEIGKRYGDSQPRDASKI 239
Db 124 KSTPEGARHFLVPSRLNKNKRYALPQSPQFKOLIMISGID---RYQIVKCFRD-EDL 178
QY 240 LVDALOKPAD-DMYSLYGNNAVVELYVKSFPDLSLRKRTILEAKQANKPASPYNLAYK 298
Db 179 RIDQPEFTQLDLEMSFATSE-----DVQISESLIKKILKEVKNFEIKEPL--LRLSYK 231

RESULT 11
MSB_HUMAN
ID MSB_HUMAN STANDARD; PRT; 1935 AA.
AC P12883; Q14904; Q16579;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
GN MYH7 OR MYHCB.
OS Homo sapiens (human).
OC Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91065634; PubMed=2249844;
RA Jaenicke T., Diederich K.W., Haas W., Schleich J., Lichter P.,
RA Pfriodt M., Bach A., Vosberg H.P.;
RT "The complete sequence of the human beta-myosin heavy chain gene and
RL a comparative analysis of its product.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90301496; PubMed=2362820;
RA Liew C.-C., Sole M.J., Yamauchi-Takahara K., Kellam B.,
RA Anderson D.H., Lin E., Liew J.;
RT "Complete sequence and organization of the human cardiac beta-myosin
RL heavy chain gene.";
RL Nucleic Acids Res. 18:3647-3651(1990).
RN [3]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=89264452; PubMed=2726733;
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT "Characterization of human cardiac myosin heavy chain genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN [4]
RP ERRATUM.
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RL Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).
RN [5]
RP SEQUENCE OF 1310-1935 FROM N.A.
RX MEDLINE=86176778; PubMed=2421254;
RA Saer L., Leinwand L.A.;
RT "Characterization of diverse forms of myosin heavy chain expressed in
RL adult human skeletal muscle.";
RL Nucleic Acids Res. 14:2951-2969(1986).
RN [6]
RP REVISIONS.
RA Leinwand L.A.;
RL Submitted (MAR-1988) to the EMBL/Genbank/DBJ databases.

```

RN [17]
 RP SEQUENCE OF 1410-1935 FROM N.A.
 RX MEDLINE-88299163; PubMed-2969919;
 RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
 RT "Molecular cloning and characterization of human cardiac alpha- and
 RT beta-form myosin heavy chain complementary DNA clones. Regulation of
 RT expression during development and pressure overload in human
 RT atrium.";
 RL J. Clin. Invest. 82:524-531(1988).
 RN [18]
 RP SEQUENCE OF 785-1935 FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;
 RX MEDLINE-90235862; PubMed-1691980;
 RA Boer E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
 RA Arnold H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RL Eur. J. Biochem. 189:55-65(1990).
 RN [19]
 RP SEQUENCE OF 1393-1935 FROM N.A.
 RX MEDLINE-87192738; PubMed-3032769;
 RA Jandreski M.A., Liew C.-C.;
 RT "Construction of a human ventricular cDNA library and
 RT characterization of a beta myosin heavy chain cDNA clone.";
 RL Hum. Genet. 76:47-53(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-96039076; PubMed-8533830;
 RA Arai S., Matsuo K., Hirayama K., Sukurai H., Tamura M., Ozawa T.,
 RA Kimura S., Imanaka S.-I., Furutani Y., Joh-O. K., Kawana M., Takao A.,
 RA Hosoda S., Momma K.;
 RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
 RT hypertrophic cardiomyopathy.";
 RL Am. J. Med. Genet. 58:267-276(1995).
 RN [11]
 RP VARIANTS CMH1 GLU-256 AND ARG-741.
 RX MEDLINE-93248216; PubMed-8483915;
 RA Panapazir L., Balakas M.C., Cyran F., Cohn G., Epstein N.D.;
 RT "Missense mutations in the beta-myosin heavy-chain gene cause central
 RT core disease in hypertrophic cardiomyopathy.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).
 RN [12]
 RP VARIANT CMH1 GLN-403.
 RX MEDLINE-90367131; PubMed-1975517;
 RA Geisterfer-Lowrance A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,
 RA McKenna W., Seidman C.E., Seidman J.G.;
 RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta
 RT cardiac myosin heavy chain gene missense mutation.";
 RL Cell 62:999-1006(1990).
 RN [13]
 RP VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.
 RX MEDLINE-92204193; PubMed-1552912;
 RA Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W.,
 RA Seidmann C.E., Seidmann J.G.;
 RT "Characteristics and prognostic implications of myosin missense
 RT mutations in familial hypertrophic cardiomyopathy.";
 RL New Engl. J. Med. 326:1108-1114(1992).
 RN [14]
 RP VARIANTS CMH1 GLN-403; CYS-453; ARG-584 AND MET-606.
 RX MEDLINE-94070863; PubMed-8250038;
 RA Watkins H., Thierfelder L., Anan R., Jarcho J., Matsuoka A.,
 RA McKenna W., Seidman J.G., Seidman C.E.;
 RT "Independent origin of identical beta cardiac myosin heavy-chain
 RT mutations in hypertrophic cardiomyopathy.";
 RL Am. J. Hum. Genet. 53:1180-1185(1993).
 RN [15]
 RP VARIANTS CMH1 GLN-403 AND VAL-908.
 RX MEDLINE-92346810; PubMed-1638703;
 RA Epstein N.D., Cohn G.M., Cyran F., Panapazir L.;
 RT "Differences in clinical expression of hypertrophic cardiomyopathy
 RT associated with two distinct mutations in the beta-myosin heavy chain
 RT gene. A 908Leu->Val mutation and a 403Arg->Gln mutation.";
 RL Circulation 86:345-352(1992).

RN [16]
 RP VARIANTS CMH1 LEU-403 AND TRP-403.
 RX MEDLINE-94075629; PubMed-8254035;
 RA Danuse E., Komada M., Feller L., Dubourg O., Dufour C., Carrier L.,
 RA Wisniewsky C., Bercevic J., Hengstenberg C., Al-Mandawi S.;
 RT "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and
 RT identification of a hot spot for mutations in the beta-myosin heavy
 RT chain gene.";
 RL J. Clin. Invest. 92:2807-2813(1993).
 RN [17]
 RP VARIANT CMH1 TRP-403.
 RX MEDLINE-94093568; PubMed-8268932;
 RA Moollman J.C., Brink P.A., Corfield V.A.;
 RT "Identification of a new missense mutation at Arg403, a Cpg mutation
 RT hotspot, in exon 13 of the beta-myosin heavy chain gene in
 RT hypertrophic cardiomyopathy.";
 RL Hum. Mol. Genet. 2:1731-1732(1993).
 RN [18]
 RP VARIANT CMH1 ASN-615.
 RX MEDLINE-93038688; PubMed-1417858;
 RA Nishi H., Kimura A., Harada H., Sasazuki T.;
 RT "Novel missense mutation in cardiac beta myosin heavy chain gene
 RT found in a Japanese patient with hypertrophic cardiomyopathy.";
 RL Biochem. Biophys. Res. Commun. 188:379-387(1992).
 RN [19]
 RP VARIANT CMH1 GLY-778.
 RX MEDLINE-93343938; PubMed-8343162;
 RA Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.;
 RT "A missense mutation of cardiac beta-myosin heavy chain gene linked
 RT to familial hypertrophic cardiomyopathy in affected Japanese
 RT families.";
 RL Biochem. Biophys. Res. Commun. 194:791-798(1993).
 RN [20]
 RP VARIANT CMH1 VAL-908.
 RX MEDLINE-93168485; PubMed-8435239;
 RA Al-Mandawi S., Chamberlain S., Cleland J., Nihoyannopoulos P.,
 RA Gilligan D., French J., Choudhury L., Williamson R., Oakley C.;
 RT "Identification of a mutation in the beta cardiac myosin heavy chain
 RT gene in a family with hypertrophic cardiomyopathy.";
 RL Br. Heart J. 69:136-141(1993).
 RN [21]
 RP VARIANT CMH1 TRP-719.
 RX MEDLINE-95179132; PubMed-7874131;
 RA Greve G., Bachinski L., Friedman D.L., Czernuszewicz G., Anan R.,
 RA Towbin J., Seidman C.E., Roberts R.;
 RT "Isolation of a de novo mutant myocardial beta MHC protein in a
 RT pedigree with hypertrophic cardiomyopathy.";
 RL Hum. Mol. Genet. 3:2073-2075(1994).
 RN [22]
 RP VARIANTS CMH1 CYS-513; ARG-716, AND TRP-719.
 RX MEDLINE-94110336; PubMed-8282798;
 RA Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S.,
 RA Vecchio C., Shono H., Nakao S., Tanaka H., Kares A., Jr., Towbin J.A.,
 RA Spirito P., Roberts R., Seidman J.G., Seidman C.E.;
 RT "Prognostic implications of novel beta cardiac myosin heavy chain gene
 RT mutations that cause familial hypertrophic cardiomyopathy.";
 RL J. Clin. Invest. 93:280-285(1994).
 RN [23]
 RP VARIANT CMH1 THR-797.
 RX MEDLINE-96047159; PubMed-7581410;
 RA Moollman J.C., Brink P.A., Corfield V.A.;
 RT "Identification of a novel Ala797Thr mutation in exon 21 of the beta-
 RT myosin heavy chain gene in hypertrophic cardiomyopathy.";
 RL Hum. Mutat. 6:197-198(1995).
 RN [24]
 RP VARIANT CMH1 CYS-453.
 RX MEDLINE-96209901; PubMed-8655135;
 RA Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C.,
 RA Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;
 RT "Malignant familial hypertrophic cardiomyopathy in a family with a
 RT 453Arg->Cys mutation in the beta-myosin heavy chain gene:
 RT coexistence of sudden death and end-stage heart failure.";
 RL Hum. Genet. 97:585-590(1996).

OS Tulipa gesneriana.
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Liliales; Liliaceae; Tulipa.
 OX NCBI_TaxID=13306;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. APELDOORN;
 RA Balk P.A., de Boer A.D.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: SUCROSE-CLEAVING ENZYME THAT PROVIDES UDP-GLUCOSE AND
 CC FRUCTOSE FOR VARIOUS METABOLIC PATHWAYS.
 CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + D-FRUCTOSE = UDP + SUCROSE.
 CC -1- SIMILARITY: TO OTHER PLANTS SUCROSE SYNTHASE..
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X96938; CAA65639.1; -
 DR InterPro: IPR000368; -
 DR InterPro: IPR001296; -
 DR Pfam: PF00534; Glycos_transf_1; 1.
 DR Pfam: PF00862; Sucrose_synth_1.
 KW Transferase; Glycosyltransferase; Multigene family.
 SQ SEQUENCE 805 AA: 92254 MW: 874989539338F0 CRC64;

Query Match 5.5%; Score 96.5; DB 1; Length 805;
 Best Local Similarity 22.9%; Pred. No. 16;
 Matches 59; Conservative 39; Mismatches 101; Indels 59; Gaps 11;
 QY 60 WPGIAYGNLFHRRATYVWYK-GVKKLALPPGSYISPLENAVPPSIDSVANSTHSLFS 118
 DB 84 WVALAI-----RPRGWVEYRVYNVELAVECS-----EYLKFKEDLYDRSSQSNFV 131
 QY 119 EETPVVQLAPSEEEVYVWGVKANSVEDLSVTLRQLRNRLFOENSYSLSPLNSLSRNN- 177
 DB 132 LE-----MDEPPNANVPRPSLSKSTGNGVQPLNRHLSTKLFPHDKESLYPL-LNPLREHNY 186
 QY 178 -EVDLFLSELQVLDHDSLSLRKH---LAKD--HSPDYSLLELAGLDETGRKYGEDS 230
 DB 187 KGTTLMLNDRLQSLASALQALRKADRYLLSTKDPYSEFNHSHFOVLGLE--KMGCDFA 243
 QY 231 EOPFDASKILVD-----ALOKPADDM-----YSLXGNNAV 260
 DB 244 SRVSENIHLLDLLEADPSTLEKLTIPWVFNVILSPHGYFAQANVLYGTPDTGGQVY 303
 QY 261 VELVYVKSFDTSILIKTR 278
 DB 304 YIIDOVRALETEMLELKIK 321

Search completed: July 3, 2001, 16:19:44
 Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:17:48 ; Search time 28.21 Seconds
(without alignments)
1641,502 Million cell updates/sec

Title: US-09-284-320-6
Perfect score: 1763
Sequence: 1 MAVFVLLALVAGVLCNEFS.....MDPGVSLITRMNOKIRMD 350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP unclassified:*
13: SP vertebrate:*
14: SP virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the local score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1759	99.8	350	4	09H2P8
2	1667	94.6	330	4	09H2P8
3	327	18.5	320	5	09H64
4	192.5	10.9	324	5	09H64
5	121.5	6.9	364	10	09S37
6	114	6.5	663	3	074359
7	109.5	6.2	408	5	022736
8	107.5	6.1	251	5	09N598
9	107	6.1	1112	10	09H47
10	105	6.0	458	5	093377
11	104	5.9	412	2	023865
12	103.5	5.9	732	2	068717
13	103.5	5.9	732	2	068717
14	103.5	5.9	925	10	09S14
15	103	5.8	470	2	09ZDN3
16	102.5	5.8	1380	5	09V108
17	102	5.8	403	4	09NXT1
18	101	5.7	412	2	092307
19	101	5.7	689	3	P87169

20	100.5	5.7	2820	11	P97526	P97526
21	99.5	5.6	861	5	09N519	09N519
22	98.5	5.6	471	3	09P723	09P723
23	98.5	5.6	728	2	056921	056921
24	98.5	5.6	1038	6	028659	028659
25	98.5	5.6	1935	4	09H1D5	09H1D5
26	98.5	5.6	2763	13	09YCV2	09YCV2
27	98	5.6	561	1	09V217	09V217
28	98	5.6	864	10	09T033	09T033
29	98	5.6	980	14	P90478	P90478
30	97.5	5.5	758	1	09U762	09U762
31	97.5	5.5	982	14	056861	056861
32	97.5	5.5	1348	5	09VAD1	09VAD1
33	96.5	5.5	533	1	09YCP2	09YCP2
34	96.5	5.5	574	10	09YV55	09YV55
35	96.5	5.5	729	2	085239	085239
36	96.5	5.5	1062	2	P95422	P95422
37	96.5	5.5	1062	2	0910Y8	0910Y8
38	96.5	5.5	1935	6	09GKR1	09GKR1
39	96	5.4	390	2	09JZF1	09JZF1
40	95.5	5.4	1020	2	09KLN4	09KLN4
41	95.5	5.4	1209	5	093620	093620
42	95.5	5.4	2471	14	09YTK3	09YTK3
43	95	5.4	2253	13	P70012	P70012
44	94.5	5.4	769	4	09H4E9	09H4E9
45	94	5.3	1778	3	09UTK4	09UTK4

ALIGNMENTS

RESULT 1
ID 09H2P8 PRELIMINARY; PRT; 350 AA.
AC 09H2P8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HT028.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HYPOPHALAMUS;
RA Yang Y., Xu X., Gao G., Xiao H., Chen Z., Han Z.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF248966; AAG44564.1; -
SQ SEQUENCE 350 AA; 39036 MW; 840A989ACC4EBD8 CRC64;

Query Match	99.8%	Score 1759;	DB 4;	Length 350;
Best local similarity	99.7%	Pred. No. 5.8e-126;		
Matches 349;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
OY	1	MAVFVLLALVAGVLCNEFSIIKSPGVFRNGNMPDGERIPDVAALSMGFSVKEDLSW 60		
DB	1	MAVFVLLALVAGVLCNEFSIIKSPGVFRNGNMPDGERIPDVAALSMGFSVKEDLSW 60		
OY	61	PELAVGNLFHPRAATYVWVKGVNKLALPGSVISYPLENAPPSLDVSANSTHLSFSE 120		
DB	61	PELAVGNLFHPRAATYVWVKGVNKLALPGSVISYPLENAPPSLDVSANSTHLSFSE 120		
OY	121	TEVVLQIAPSEERYVWVKANSVFDLSVTLRLRNRLFOENSIVLSPNLSIRNEVD 180		
DB	121	TEVVLQIAPSEERYVWVKANSVFDLSVTLRLRNRLFOENSIVLSPNLSIRNEVD 180		
OY	181	LEFLSELQVLDHDSLSLRHKLAKDSPLDYSLELAGLDEIGKRYGDESDQPRASKTL 240		
DB	181	LEFLSELQVLDHDSLSLRHKLAKDSPLDYSLELAGLDEIGKRYGDESDQPRASKTL 240		
OY	241	VDALOKRADMTSLYGGNAVELYVKSFTDSLIRKRTTILEAKQAKNPASPYNLATKYKN 300		

Db 241 VDLQKRFADMDYSLYSGNAVVELVYVSKSFDTSLRRTTLEAKKANKPASPYNLAAYKN 300
 QY 301 FEYSVFNMMVIMIALALAVITTSYNIMNMDPGYSIIYRMTNOKIRMD 350
 Db 301 FEYSVFNMMVIMIALALAVITTSYNIMNMDPGYSIIYRMTNOKIRMD 350

RESULT 2
 ID 09UG89 PRELIMINARY: PRT: 330 AA.
 AC 09UG89;
 DT 01-MAY-2000 (TReMBLrel. 13. Created)
 DT 01-MAY-2000 (TReMBLrel. 13. Last sequence update)
 DE 01-MAY-2000 (TReMBLrel. 13. Last annotation update)
 DE HYPOTHEICAL 37.0 KDA PROTEIN (FRAGMENT).
 GN DKF2P56400582.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN.
 RA Poustka A., Klein M., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL049929; CAB43210.1; -.
 FT Hypothetical protein.
 KM NON_TER 1
 SQ SEQUENCE 330 AA; 37004 MW; 58098DCDF5B6CFE2 CRC64;

Query Match 94.6%; Score 1667; DB 4; Length 330;
 Best Local Similarity 99.7%; Pred. No. 5,3e-119;
 Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 21 ILSKSGSVFRNCGNWPPIPERIPDPAALSMGFSYKEDLSMPGLAVGNLFRPRATYMWV 80
 Db 1 ILSKSGSVFRNCGNWPPIPERIPDPAALSMGFSYKEDLSMPGLAVGNLFRPRATYMWV 60
 QY 81 KGVNKLALPGSVSYSTPLENAPVPSLDSVANSISLSFSEEPVVLQAPSEERVYWGKA 140
 Db 61 KGVNKLALPGSVSYSTPLENAPVPSLDSVANSISLSFSEEPVVLQAPSEERVYWGKA 120
 QY 141 NSVFEDLSVTLRDLRRRLRPOESVYSSPLNSISRRNVEDLLFSELQVYHDISSLSRH 200
 Db 121 NSVFEDLSVTLRDLRRRLRPOESVYSSPLNSISRRNVEDLLFSELQVYHDISSLSRH 180
 QY 201 KHLAKDHSPLDYSLELAGLDEIGKRYGDESEOPRDAKTIYDALQKPADDMYSLYGNAV 260
 Db 181 KHLAKDHSPLDYSLELAGLDEIGKRYGDESEOPRDAKTIYDALQKPADDMYSLYGNAV 240
 QY 261 VELVYVASFDTSLRRTTLEAKOANKPASPYNLAAYKNFEYSVFNMMVIMIALALA 320
 Db 241 VELVYVASFDTSLRRTTLEAKOANKPASPYNLAAYKNFEYSVFNMMVIMIALALA 300
 QY 321 VLTTSYNIMNMDPGYSIIYRMTNOKIRMD 350
 Db 301 VLTTSYNIMNMDPGYSIIYRMTNOKIRMD 350

RESULT 3
 ID 09VHG4 PRELIMINARY: PRT: 320 AA.
 AC 09VHG4;
 DT 01-MAY-2000 (TReMBLrel. 13. Created)
 DT 01-MAY-2000 (TReMBLrel. 13. Last sequence update)
 DE 01-MAY-2000 (TReMBLrel. 13. Last annotation update)
 GN CG8444.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephyroidae; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Morten J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abill J.F., Agayari A., An H.-J., Andrews-Plankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reiert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spedding A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye Y., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhang L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003682; AAF54350.1; -.
 DR Flybase: FBgn0037671; CG8444.
 SQ SEQUENCE 320 AA; 34421 MW; 401BCBAAA4FBA7E5 CRC64;

Query Match 18.5%; Score 327; DB 5; Length 320;
 Best Local Similarity 26.3%; Pred. No. 4.9e-17;
 Matches 94; Conservative 74; Mismatches 142; Indels 48; Gaps 11;

QY 1 MAVVNVLLALVAGV-LGNFESILKSGSVFRNCGNWPPIPERIPDPAALSMGFSYKEDLS 59
 Db 2 LRVFVIFSLFIALINASGEFTVLRPKALSEK-GNDALSHYGVADVLSMGNAVSGGTN 60
 QY 60 WPGVAVNLFRRFRATVWVWVGKYNKLALPGSVSYSTPLENAPVPSLDSVANSISLSFSE 119
 Db 61 WNGGLINDPENLAKGYLLVHVGIGHYT-TAGNVKTYELTGS--GIDASLAL----- 110
 QY 120 ETPVVDLALASERVYVWVGKANSVFEDLSVTLRDLRRRLRPOESVYSSPLNSISRRNEV 179
 Db 111 -----AAELEANAPYCDIN-----FEQFDGVQANKS-CFQDEAPAPKPTHTLWPS-- 157
 QY 180 DLFLESLQVYHDISSLSRKHHLAKDHSPLDYSLELAGLDEIGKRYGDESEOPRDAKSI 239
 Db 158 -LHTADKQFLQEVGFTNSADHLAEKAPSVWMLRLRVSYDVGAAKHAKGSAVAPEANKL 215
 QY 240 LVYDALQKPADDMYSLYGNAVVELVYVSKSFDTSLRRTTLEAKOAK-----NPASP 292
 Db 216 LSAAISRLT-----AASOKSSDVSILFVQPTTEKDVAAASRAKRDITIAASTNP 261


```

RESULT 6
ID 074359 PRELIMINARY; PRT; 663 AA.
AC 074359.
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHEtical 75.5 KDa PROTEIN.
GN SPEC28E12.02 OR SPACB6.13.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-.
RA Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031517; CA20649.1; -.
DR EMBL; AL049769; CAB42374.1; -.
DR InterPro; IPR000566; -.
DR InterPro; IPR000958; -.
DR Pfam; PF00013; KH-domain; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR SMART; SM00322; KH; 1.
KW Hypothetical protein.
SQ SEQUENCE 663 AA; 75492 MW; 6898B18AC142A8BA CRC64;

Query Match 6.5%; Score 114; DB 3; Length 663;
Best Local Similarity 21.5%; Pred. No. 2.5;
Matches 62; Conservative 43; Mismatches 92; Indels 92; Gaps 13;

QY 13 GVLGNF-STLSPGVSFVRNGNMPICERIPDVAALSMGYSKEDLSMPGLAVGNLFR 71
DB 414 GVSSESLSTSVSTGSIETNG-----IGEK-----MSFSLKLKLSIP----- 451
QY 72 P-----RAYWVWVGYNKALPPGSVISPLENAVPSLDSVANSIHSLSFEETPVYIQ 126
DB 452 PTEVVAOIAICASGVEML-LKTNGIEYFGQENTVPIAMKASKIFPKFGOSQWQTL 510
QY 127 LAPSEEVYWGKANSVEDLSVTLR-QLRNRLFOENSIVLSLPLNLSRNNEVDLFLS 185
DB 511 EAPKDDDFISGKKKGLDKYKQOCRENLKNC-----DILFCP 548
QY 186 ELGYLHDISSLSRKHKLAKHSPDLYSLEAGL-----DELGR 225
DB 549 QSTISITV-----DIISDELREVIKGNITMLLEPPAEHMYVPELHK 592
QY 226 -YGDSEQFRDASKILVDALOKFADWYSLYSGNAVVELVTVNSPFTSL 273
DB 593 LIGRGEIOIRVTK-LVNSYIEFTPTFCYGHVLLR--TPSKFSENL 638

RESULT 7
ID 022736 PRELIMINARY; PRT; 408 AA.
AC 022736.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE T24D5.2 PROTEIN.
GN T24D5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wilkinson J.;

```

```

RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Harkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldmann P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z68012; CAA92020.1; -.
SQ SEQUENCE 408 AA; 46861 MW; B0B171CA84197EB9 CRC64;

Query Match 6.2%; Score 109.5; DB 5; Length 408;
Best Local Similarity 20.8%; Pred. No. 2.6;
Matches 43; Conservative 47; Mismatches 84; Indels 33; Gaps 7;

QY 81 KGVNK-LALPPGSVTS-----YPLENAVPSLDSVANSIHSLSFEETPVYQLAPSEE 132
DB 103 KAMNPLKGFGEVLAKYFQFCHELNNTYKKNKTIIRLDAYKMTFSEKDSILIRSDGS 162
QY 133 RYVWGKANSFEDSVTLRQLRNRLFOENSIVLSLPLNLSRNNEVDLFLSELOVLD 192
DB 163 TVHMGISNNTLFRNNMAIEELRNMLSQNIRYKSLFVTRNSQ-----MGENDLSKC 215
QY 193 ISSLSHKKHLAKHSPDLYSLEAGLDITGK-----RYGE--DSEQFRDASKILVDA 243
DB 216 VSTTANLTVYMDSPFNIEIVKLRHPLNKFILMCSGEYGNIMELQVRAANLKLGD 275
QY 244 LQKFADWYSLYSGNAVVELVTVNSPD 270
DB 276 LSKITDQ-----LNSITATNFE 293

RESULT 8
ID 09N598 PRELIMINARY; PRT; 251 AA.
AC 09N598.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE Y110A7A.11.
GN Y110A7A.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Ozerky P., Wohldmann P., Carter T.;
RT "The sequence of C. elegans cosmid Y110A7A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;

```

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC006708; AAF60428.1; -
 SO SEQUENCE 251 AA; 28808 MW; 3E19F7FDD9B3FE40 CRC64;

Query Match 6.1%; Score 107.5; DB 5; Length 251;
 Best Local Similarity 22.1%; Pred. No. 1.7;
 Matches 66; Conservative 48; Mismatches 86; Indels 99; Gaps 15;

QY 55 KEDLSMGLAVNGLTHRRPRATVYMWGVKAKLALPPGSVISYPLENNAVPSLDSVANSIH 114
 DB 23 KEDMS-----ANVSATKILKKMI-----YTLQN-----DETNDIL 54
 QY 115 SLFSEETPVYLOL-----APSERVYMWGKANSYFEDLSVTLKQLRNRLFOEN--- 162
 DB 55 HEYHQE---VLQKLLABAESKSSAEERLKVIEKIPRVFPQVYTVADSKNNNEAESYDF 111
 QY 163 SVLSLPLNSLRNNEDVLLFLSELQVLDHDISSLSRRKHLAKHSDIXSLFAGLDEI 222
 DB 112 EKESAGLRATQRS-----YRSOLR-----KQLSSKRHAQDTS----- 147
 QY 223 GKRGESEQFRD--ASKILVDALOKFADMYSLY--GNAVAVELTVKSFPTSLIRKTR 278
 DB 148 -----EDGEFMKMLVEBELANSLATWASFKTMSAGDIKE-----DTE-----R 190
 QY 279 TILKAKAKNPASPYNL-----AKYFEXSVFNMVLMIALALAVIITSYNIW 329
 DB 191 AILMAKEVDNKTALGIOSEVERERHAYKGYD--CFKVMILVLIEMSFVSVLMKKIF 246

RESULT 9
 Q9FH47 PRELIMINARY: PRT; 1112 AA.
 AC Q9FH47;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE GBAIAD5516.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLOMBIA;
 RX MEDLINE=20181125; PubMed=10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
 RA Tabata S.,
 RA "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
 RA features of the regions of 3,076,755 bp covered by sixty P1 and TAC
 RA clones."
 RL DNA Res. 7:31-63(2000).
 DR EMBL: AB022212; BAB08882.1;
 SO SEQUENCE 1112 AA; 127094 MW; 7C11A579BB26DE85 CRC64;

Query Match 6.1%; Score 107; DB 10; Length 1112;
 Best Local Similarity 21.5%; Pred. No. 19;
 Matches 74; Conservative 62; Mismatches 114; Indels 94; Gaps 15;

QY 75 TWYMWGVKVKL---ALPPGSVISYPLENNAVP-----SLDSVANSIHSLFSEETPVV 124
 DB 592 TTKLFVAVADVEGALHP-FLDPLDEDCLNFIHSHKSSLDSDKAVLOAIDLKSVYTKQV 650
 QY 125 LQAPSERVYMWGKAN-----SVEDLSVTLKQLRNRLFOENSYSLSPLNSLS 174
 DB 651 LLM---PTKILLDNRISLNNITRLSAPDNRTYLLQLRPLFLN---IYMSKSKS 704
 QY 175 RNNEVDLFLSELQVH-----DISSLSRHLKHLADH 207
 DB 705 DAAEADLLNELEKEKIQPREKPKSKKPKRRORTKKKRPDSTISSLLDKYEHKPEST 764

QY 208 SPDLYSLE-----LAGLDEIGKRGEDSEQFROASKILVDALOKFADMYSLYGN 258
 DB 765 SPSLRYEEDSEHPEDALASFTDMHSHREDP-----LSDHLESAGAAAR--N 813
 QY 259 AVE-----LVYKSFDTSLIRKTRITLEBAKQAKNPASPYNL-----AKYFEXSVV 306
 DB 814 SALDMLTKALLSLIKFKEDLNQSOQIODOBOQVPTALQNLFTAVASPAIDEGVYSL 873
 QY 307 FNMVL-----MIMIALAVIITSYNIW--MDPGYDSTIYEM 342
 DB 874 LSDFLYSQEFSLSSSDAKAVAVIYLMRCMKNPRESLVTRL 917

RESULT 10
 Q93377 PRELIMINARY: PRT; 458 AA.
 AC Q93377;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE C44H4.1 PROTEIN.
 GN C44H4.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;
 OC Rhabdilitidae; Pelodertinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smye R.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Crexton A., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
 RA Watson A., Welnsstock L., Wilkinson-Spoat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL Nature 368:32-38(1994).
 DR EMBL: Z79598; CAB01867.1;
 DR InterPro: IPR001611;
 DR InterPro: IPR003591;
 DR Pfam: PF00560; LRR: 10.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SM00369; LRR_TYP.1.
 SO SEQUENCE 458 AA; 51171 MW; 8A6810B119684152 CRC64;

Query Match 6.0%; Score 105; DB 5; Length 458;
 Best Local Similarity 22.5%; Pred. No. 6.7;
 Matches 69; Conservative 47; Mismatches 116; Indels 74; Gaps 12;

QY 18 EFSILKSPGSVFRNQNPIGERIPDVAALSKGSFVXEDLSWGLAVNGLFHRPRATVM 77
 DB 149 QKTLISLPDMLI-----EYQDNAPLS-----YHSRDSLK 179
 QY 78 VYMWGVKAKLALPPGSVISYPLENNAVPSLD--SVANSIHSLFSEETPVYLOLAPSERVYM 136
 DB 180 LDLSNNMLTAIHTPTGLG--LENLSQSLDKNLSLSPQALENIPSLDLSGVNRLHT 237
 QY 137 VGRANV-----FEDLSVTLKQLR--NRLFOENSYSLSPL--NSLSRNNEVDLFLSEL 187
 DB 238 ISR-NSLPLPNLKSLSLEVNQIRLIPDSFSEFPLSLYLVGNLLITSIDASFHLIGL 296
 QY 188 QVL-----HSSLSLRHK-----HLAKHSPDLYSLELGLDEIGKRGEDSE 231

Db 297 KVLSSNNKDIITSIQANGKLSVOOFKLCIILAFHAPSLIRLELFC----- 344
QY 232 QFRDASKIIDLQKADDMWSLYGNAVVELVWYKPTSLIKRTILIEAKQKNPAS 291
Db 345 ---SISRIEPEKSLQKQYHIOYILLSRNOITQINAVDDAFSOL-PMILSLIDLSNRLSL 400
QY 292 PYNLAY 297
Db 401 PSNVIY 406
RESULT 11
025885
ID 025885 PRELIMINARY; PRT; 412 AA.
AC 025885;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 47.1 KDA PROTEIN.
GN HPI327.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
RA Flatschmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush D., Zhou L., Kirness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodok A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.R.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
DR EMBL: AE000634; AAD08377.1; -;
DR TIGR: HPI327; -;
DR InterPro: IPR003423; -;
DR Pfam: PF02321; OEP; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 47115 MW; 17240849BRA15DD CRC64;

Query Match 5.9%; Score 104; DB 2; Length 412;
Best Local Similarity 22.0%; Pred. No. 6.8;
Matches 70; Conservative 55; Mismatches 101; Indels 92; Gaps 15;

QY 48 LSMGFYSKEDLSMPGLAVGNLFH-RRRAIVYVWVGKVALPFGSVISYLEN----- 100
Db 97 MSLSLSKQDLNMGKLTOSQMDIEKOKKILEKTKQOLAI--SLMNGENKNOOE 153
QY 101 --AVPESLDSVANSIHSLFSEETPVVLOLAPSE-RRVYVWGKANSVEFDLSTLRQLNR 157
Db 154 IELTKTAIKKLENTLYQANHSSSPNIIAIAKLEIKLSQLEIKKNLLEALSMASHSMEL 213
QY 158 LFOENSVLSLPLNSLRNNNEVDLFLSELQYLDHISLSRKHAKLAKHSDLYSLELA 217
Db 214 AFKNELSLSTAKRN-FERNR------QELNHISA-----TNYDIAIA 249
QY 218 GLDEIGKRYGDESDQFRDASK--TLVD-----ALQKFADDMS-----LYGG 257
Db 250 RLDE-----EKSQMDITLAKSKSLSDVWVTGVYFRSKQYVNDMFIALSILPITGX 303
QY 258 NAYVELVYKASDPT-----SLIKTRTIT--LEA-----KOAKNPA 290
Db 304 QA-KLVEQKKESLVEFKSEVENTKTKTHLALDKLKLKLTLOKLNESINKITIKONEKIA 361
QY 291 SPYNLAYVYKFEYSVFN 308

Db 362 QIVALDLKSNQGYNAVYN 379

RESULT 12
068717
ID 068717 PRELIMINARY; PRT; 732 AA.
AC 068717;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PROTEIN KINASE A HOMOLOG.
GN YPKA.
OS Yersinia pestis.
OG Plasmid pCDL.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM;
RX Hu P., Elliott J., McCready P., Skowronski E., Barnes J.,
RA Kobayashi A., Carrano A.V., Brubaker R., Garcia E.,
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5;
RX MEDLINE=98427122; PubMed=9746557;
RA Perry R.D., Straley S.C., Fetherston J.D., Rose D.J., Gregor J.,
RA Blather F.R.;
RT "DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1 of
RT Yersinia pestis KIM5."
CC Infect. Immun. 66:4611-4623(1998).
CC -I- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: AF053946; AAC62602.1; -;
DR EMBL: AF074612; AAC69765.1; -;
DR InterPro: IPR000719; -;
DR InterPro: IPR002290; -;
DR InterPro: IPR003547; -;
DR PRINTS: PR01373; YERKSTKINASE.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR SMART: SM00220; S_TKC; 1.
KW ATP-binding; kinase; Plasmid; Serine/threonine-protein kinase;
KW Transferrase.
SQ SEQUENCE 732 AA; 81732 MW; 1ECB0099EE2D3430 CRC64;

Query Match 5.9%; Score 103.5; DB 2; Length 732;
Best Local Similarity 21.2%; Pred. No. 18;
Matches 73; Conservative 51; Mismatches 103; Indels 117; Gaps 17;

QY 12 AGVLGNEFSILKPGSVFNRGNMPLIGERIPDVAALSM-----GFSYKEDLSMPG 62
Db 264 AGVYHNDI-----KPGNVVFDRA-----GE--PVVIDLGLHSRSGEQPKFT--ESFKAPE 311
QY 63 LAGNLFHRRATVWVY-----KGVNKLALPPGSVISTYLENAVP 103
Db 312 LGVNGNLGASERKDVIVYVSTLLHCIGFEPKNEIKPNQIGLRTTSPRAHYD---ENGIV 368
QY 104 FSLDSVANSIHSLFSEETPVVLOLA-----PSEERYVWVGKANSVEFDLSTLRQLNR 157
Db 369 IHRGGLA-GVETAVYRFTDILGVSNADSRPDSNENARLHFLSDGTDEBSAKQI--LKDT 425
QY 158 LFOENSVLSLPLNSLRNNNEVDLFLSELQYLDHISLSRKHAKLAKHSDPD----- 211
Db 426 LTGEMKPLST-----DYKRIIP-KKLRELSDILKRTLSSAATKQLDGCVLSD 472
QY 212 YSLELAGLDEIGKRYGDESDQFRDASKTLVVALQKADDMWSLYGNAVVELVYKSPDT 271
Db 473 LDMVALDLKAREGVDKQD-----LKSFN----- 498
QY 272 SLIKTRTITL-----AKQAKNPASPYNLAYVYKFEYSV 306

Db 499 SLIKTYRVIEDYVKGREGDTKNSSTEVSPY---HRSNFMLSIV 539

```

RESULT 13
ID 09R112 PRELIMINARY; PRT; 732 AA.
AC 09R112;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PUTATIVE TARGETED EFFECTOR PROTEIN KINASE.
GN YPCD1.72c.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_Taxid=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 BIOVAR ORIENTALIS;
RA Baker S.G., Mungall K.;
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 BIOVAR ORIENTALIS;
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 BIOVAR ORIENTALIS;
RA Karlyshev A.V., Wren B.W.;
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: A117189; CAB5494.1; -.
DR Interpro: IPR000719; -.
DR Interpro: IPR002290; -.
DR Interpro: IPR003547; -.
DR PRINTS: PR01373; YERSTKINASE.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM. 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST. 1.
DR SMART: SM00220; S_TKc. 1.
KW ATP-binding; Kinase; Plasmid; Serine/threonine-protein kinase;
KW Transferase.
SQ SEQUENCE 732 AA; 81760 MW; DEF803AAE2E5BCD CRC64;

Query Match 5.9%; Score 103.5; DB 2; Length 732;
Best Local Similarity 21.2%; Pred. No. 18;
Matches 73; Conservative 51; Mismatches 103; Indels 117; Gaps 17;

QY 12 AGVLGNEFSLKSPGVVFRNGMPTIGERIPDVAALSM-----GFSVKEDLSMPG 62
DB 264 AGVNHDI-----KGNVVFDRAS---GE--PVYIDGLSRSGEQPKQFT--ESFAPE 311
QY 63 LAVGNLFHRPRATYVWV-----KGNKALALPGSGVSYLLENAAV 103
DB 312 LGVGNLGNASEKSDVFLVYVSTLLHLCIGEPKNEIKPQGLRFTTSPAHVMD---ENGTP 368
QY 104 FSLDSVANSIHSLSSEETPVYQLA-----PSEERYVWGNKANSVFEDLSVTLROLNR 157
DB 369 IHRPGIA-GVETAVTREFITILIGVSADSRHPSNEARHFEFLSGTIDERSAKQI--LKD 425
QY 158 LFOENSVLSLPLNLSRNNEVDLFLSELQVLDHDSLSLRKHLADHSPDL----- 211
DB 426 LTGENSPILST-----DYKRITP-KKLELSDLRLTLSSAATKQLDGCVLSD 472
QY 212 YSELGLDELIGKRGEDSEQFPDASKIIVDALOKFADDMYSLYGGNAVELVTYKSPDT 271
DB 473 LDTMLVALDKAEREGVDKQD-----LKSFN- 498
QY 272 SLIKRTTILE-----AKQAKNPASPYNLAYKXNYFEYSV 306

```

Db 499 SLIKTYRVIEDYVKGREGDTKNSSTEVSPY---HRSNFMLSIV 539

```

RESULT 14
ID 09S1X4 PRELIMINARY; PRT; 925 AA.
AC 09S1X4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PUTATIVE LRR RECEPTOR PROTEIN KINASE.
GN AT2G16250.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ranning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL: AC007047; AAD2312.1; -.
DR Interpro: IPR000719; -.
DR Interpro: IPR001611; -.
DR Interpro: IPR003592; -.
DR Pfam: PF00069; pkinase. 1.
DR Pfam: PF00560; LRR. 7.
DR PRINTS: PR00019; LEURICHRPT.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM. 1.
DR SMART: SM00370; LRR. 1.
KW ATP-binding; Kinase; Receptor; Transferase.
SQ SEQUENCE 925 AA; 101125 MW; C363800D4ABC39D3 CRC64;

Query Match 5.9%; Score 103.5; DB 10; Length 925;
Best Local Similarity 22.6%; Pred. No. 26;
Matches 86; Conservative 43; Mismatches 126; Indels 125; Gaps 16;

QY 4 FVYLL-----ALVAGVLGNEFS---ILKSPGVVFRNGMPTIGERIPD----- 44
DB 11 FVLLLCVLVFPDCVVGQTSRFSKLLILNLRSSLGRLGTDPWKGDPCVDMRGIOCE 70
QY 45 -----VAALSMGFSV-----KEDLSWGLAVGNLFHRPRAT 75
DB 71 NGSILGINISGRFRTKGLKLNPFQSVDPDLRLTLSTFNASGLALPG--TIRENF----- 123
QY 76 VWMVWGVNKKALALPGSVISYPLENAVPPSLDSVA-----NSHSLSFSEETPVYQL 126
DB 124 -----GSLALALEVLDLSSCVNGVVPFTIGNLTLSTLNLSONSLTSLVPSLSGLQLN 177
QY 127 LAPSE-ERYVWGNKANSVFEDLS--VTLROLNRNL-----POENSVL 165
DB 178 LQDLDSLRNSFTGVLPSPFSLKNLTLTDVSNVLTGP IIPGLGALSKLHLNFSNSNPS 237
QY 166 SLSLP-----INSLSRNNEVDLFLSELQVLDHDSLSLRKHLAKDHPD 210
DB 238 SPIPSELGLVNVDPFLSLINSLSGSVQDELRLKSLKLDLMAIGDMLLGLTPV-----D 291
QY 211 LVSLLELAGDELIGKRGEDSEQFPDAS-----KTIIVDALOKFADDM--YSLYGGNAVEL 263
DB 292 LFSAE-SQGLTVIVARENGFGSLPVCWSLPKRLIIDIAKNNFTGLLTYSSYSDQIAEM 350
QY 264 VTYKSPD-----TSLIKRTR 278

```

DB 351 VDISNTFYGELTFLRRFR 370

RESULT 15

ID Q92DN3 PRELIMINARY; PRT: 470 AA.

AC Q92DN3;

DT 01-MAY-1999 (TREMUREL. 10, Created)

DT 01-MAY-1999 (TREMUREL. 10, Last sequence update)

DT 01-MAY-2000 (TREMUREL. 13, Last annotation update)

DE GRANOSINE PENTAPHOSPHATE PHOSPHOHYDROLASE (GPPA).

GN R294.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=782;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-MADRID E.

RX MEDLINE-99039499; PubMed-9823893;

RA Andersson S.G.E., Zomorodipour A., Andersson J.O., Naeslund A.K.,

RA Sierhertz-Ponten T., Alsmark U.C.M., Podowski R.M.,

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of

RT mitochondria."

RT Nature 396:133-140(1998).

RU EMBL: AJ235271; CAA14755.1;

SQ SEQUENCE 470 AA; 53780 MW; 74F7E97599D02903 CRC64;

Query Match 5.8%; Score 103; DB 2; Length 470;

Best Local Similarity 21.7%; Pred. No. 10;

Matches 65; Conservative 56; Mismatches 123; Indels 56; Gaps 12;

QY 45 VAALSMGFSVKREDLSMRGLAVGNLFRRPATYMWVGVNKL-ALPRGSVSYPLENAPV 103
DB 117 ISGISDAFGIVADLGGSLAHYV-----KVGRLKSLPLGTKI----- 157
QY 104 FSLDSVANS-----IHSLFSEPTPVVLQLAPESEERYVWYGA-----NSVFEDLSVTL 151
DB 158 -----IANSFSDVGLITKMLEEFGVA-----HYPNLYLIGGALRLMSRIYMESINVP 207
QY 152 RQLRNRLFOENSIVLSLPLNSLRNNEVDLLFLSLQYLDHDSLSLRHKHAKDHPDL 211
DB 208 KNLNHN--FEINRVEPELTELKLSIDKLSTYEDOKAINTNAVLT--KAMLVFSPK 262
QY 212 YSLELAGLDEIGKRYGDESEOFADASKIL---VDALQKPADMYSLGNAVVELVTVK 267
DB 263 IISNNGLKE-GVRF--DSLPHYTEKDIYERVKRLVNFDRNICKIEKYIFALQYLLIN 319
QY 268 SFQTSILRTKITLAKAKN--PASPYLAKYKVFESYVFNWVIMIMIALAVITTS 325
DB 320 SDATYTLITELAIMLAQYNKNIDKTLRANVSEFLLSSDIPFSHRQDMGTALTYTYTA 379

Search completed: July 3, 2001, 16:20:24
Job time: 156 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 16:16:08 ; Search time 16.87 Seconds

(without alignments)
417,938 Million cell updates/sec

Title: US-09-284-320-6

Perfect score: 1763

Sequence: 1 MAFFVLLALVAGVIGNEFS.....MDPGYSIIYRMNOKIRMD 350

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	441	25.0	93	4	US-08-905-223-402
2	100.5	5.7	2020	1	US-07-551-531-2
3	100.5	5.7	2485	5	PCT-US94-00198-1
4	100.5	5.7	2485	5	PCT-US94-00198-2
5	100.5	5.7	2818	1	US-08-510-284-1
6	100.5	5.7	2818	1	US-08-411-389-2
7	100.5	5.7	2818	2	US-08-449-933-2
8	100.5	5.7	2818	4	US-07-966-049A-2
9	93	5.3	2627	2	US-08-751-189-3
10	93	5.3	2627	2	US-09-060-836-3
11	93	5.3	2627	4	US-09-184-445-3
12	91	5.2	448	4	US-09-143-954-4
13	91	5.2	457	4	US-09-143-954-2
14	91	5.2	3066	4	US-09-952-127-12
15	90.5	5.1	673	4	US-09-063-950-2
16	90.5	5.1	2987	2	US-08-970-269A-29
17	90.5	5.1	3959	2	US-08-970-269A-30
18	89	5.0	903	2	US-08-209-521-24
19	89	5.0	903	4	US-08-961-810-134
20	89	5.0	903	4	US-08-352-902D-134
21	88.5	5.0	790	4	US-08-960-780-4
22	88.5	5.0	790	4	US-09-073-898-4
23	88.5	5.0	1663	3	US-08-793-126-1
24	88.5	5.0	1663	4	US-09-132-271-1
25	88	5.0	473	5	PCT-US93-08386-9
26	88	5.0	577	1	US-08-484-105-24
27	88	5.0	577	1	US-08-484-106-24

28	87.5	5.0	789	1	US-08-471-033-29	Sequence 29, Appl
29	87.5	5.0	789	2	US-08-471-044-29	Sequence 29, Appl
30	87.5	5.0	789	2	US-08-463-483A-29	Sequence 29, Appl
31	87.5	5.0	789	2	US-08-471-046A-29	Sequence 29, Appl
32	87.5	5.0	789	2	US-08-470-566B-29	Sequence 29, Appl
33	87.5	5.0	789	2	US-08-838-219B-2	Sequence 29, Appl
34	87.5	5.0	789	2	US-08-469-334A-29	Sequence 29, Appl
35	87.5	5.0	789	3	US-09-300-529-29	Sequence 29, Appl
36	87.5	5.0	789	3	US-09-233-336A-2	Sequence 29, Appl
37	87.5	5.0	789	4	US-09-233-752A-2	Sequence 2, Appl
38	87.5	5.0	790	4	US-08-960-780-8	Sequence 2, Appl
39	87.5	5.0	790	4	US-09-073-898-8	Sequence 8, Appl
40	87.5	5.0	906	1	US-08-220-151-9	Sequence 8, Appl
41	87.5	5.0	906	1	US-08-413-118-9	Sequence 9, Appl
42	87.5	5.0	906	3	US-08-473-446-9	Sequence 9, Appl
43	87	4.9	336	3	US-09-095-163-2	Sequence 2, Appl
44	86.5	4.9	509	2	US-08-558-505-2	Sequence 2, Appl
45	86.5	4.9	509	2	US-08-890-980-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-905-223-402
Sequence 402, Application US/08905223
Patent No. 6222029
GENERAL INFORMATION:
APPLICANT: Edwards, Jean-Baptiste D.
APPLICANT: Duclert, Aymeric
APPLICANT: Lacroix, Bruno
TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
NUMBER OF SEQUENCES: 503
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 402:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: sig_peptide
LOCATION: -16...-1
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 12.5
OTHER INFORMATION: seq FVLLALVAGVIG/NE
US-08-905-223-402

Query Match 25.0%; Score 441; DB 4; Length 93;
Best Local Similarity 95.6%; Pred. No. 3.9e-38;
Matches 86; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVFVLLALVAGVILGNEFSLKSPGVFRNGMNPICRIPDVAAALSMGFSVKEDLSW 60
DB 1 MAVFVLLALVAGVILGNEFSLKSPGVFRNGMNPICRIPDVAAALSMGFSVKEDLSW 60

QY 61 PGLAVGNLFHPRPRATVVMVKGYNKLALPG 90
DB 61 PGLAVGNLFHPRPRATVVMVKGYNKLALPG 90

RESULT 2
US-07-551-531-2
; Sequence 2, Application US/07551531
; Patent No. 5227292

GENERAL INFORMATION:
APPLICANT: WHITE, Raymond L.
APPLICANT: O'CONNELL, Peter
APPLICANT: VISKOCHEL, David H.
APPLICANT: CAMTHON, Richard M.
TITLE OF INVENTION: NEUROFIBROMATOSIS TYPE 1 GENE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/551,531
FILING DATE: 19900712
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: IHEN, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 19780-94649
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2020 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-551-531-2

Query Match 5.7%; Score 100.5; DB 1; Length 2020;
Best Local Similarity 22.2%; Pred. No. 0.78;
Matches 86; Conservative 60; Mismatches 140; Indels 101; Gaps 18;

QY 13 GVIGNEFSLKSPGVFRNGMNPICRIPDVAAALSMGFSVKEDLSWPGIAY 65
DB 1302 GGIGSIRAKVADVAALVAGVILGNEFSLKSPGVFRNGMNPICRIPDVAAALSMGFSVKEDLSW 1261

QY 66 GNLFHPRATVVMVKGYNKLALPG-----SVISYLENAVPSIDS-----VANSIHS 115
DB 1262 -----LARYMLMSTFNSGLDVAALHPLYLHYVTF-LVATGPLSLASTGIVINIIHS 1313

QY 116 L-----FSECTPRVYQLAPSE---ERYVMGKANSYFEDLSVTLR-QLNRRLPEQNSV 164
DB 1314 LCTGSOJHFESECTPRVYQLAPSE---ERYVMGKANSYFEDLSVTLR-QLNRRLPEQNSV 1373

QY 165 -LSLPLUNSLSNNEVDLLPLSELQVLDHISL--LSRHKLAK-----DHSPLYSLELA 217

DB 1374 ERFTALSTLETVEA-LEIME-ACMRDIPCTCKWDQWTELQAFRAQVNPSPLOPRALV 1431
QY 218 GIDETGKR--YGEDSEQFRDASKIIVDLQKFADMTS----- 253
DB 1432 VFQCSIKRSHSGOIKQIRLSKALESCIK--GPDYNSQVLEAVYIALTKLOPLKND 1489

QY 254 -----LYGNNAVEL--VTWKSFTSLIKRTILEAKQAKNPASP----- 292
DB 1490 SPLHAKLVAVAVALLQIDENVNLXSAGTALLQONLHTLDSIRTFNOKSPFEVFAIRNPLE 1549

QY 293 -----YLAATKYNFEYSVFNKV 310
DB 1550 WHCKQMDHFGVGINFNSNFALVGHLL 1576

RESULT 3
PCT-US94-00198-1
; Sequence 1, Application PC/TUS9400198

GENERAL INFORMATION:
APPLICANT: Schering Corp.
TITLE OF INVENTION: RAS Associated GAP Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering Corp.
STREET: 1 Giraldo Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 94304-1104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 6.0.8
SOFTWARE: Microsoft Word 5.1a

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00198
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/004,824
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Iuno, Paul G.
REGISTRATION NUMBER: 32,743
REFERENCE/DOCKET NUMBER: DX0352 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)822-7255
TELEFAX: (201)822-7039

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2485 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 564..9380
PCT-US94-00198-1

Query Match 5.7%; Score 100.5; DB 5; Length 2485;
Best Local Similarity 22.2%; Pred. No. 1.1;
Matches 86; Conservative 60; Mismatches 140; Indels 101; Gaps 18;

QY 13 GVIGNEFSLKSPGVFRNGMNPICRIPDVAAALSMGFSVKEDLSWPGIAY 65
DB 1667 GGIGSIRAKVADVAALVAGVILGNEFSLKSPGVFRNGMNPICRIPDVAAALSMGFSVKEDLSW 1726

QY 66 GNLFHPRATVVMVKGYNKLALPG-----SVISYLENAVPSIDS-----VANSIHS 115

```

Db 1727 -----LARYMLMSTFNNSLDVAHLPLYLHVYTF-LVATGPLSLRSTGVLINITHS 1778
QY 116 L-----FSEETPVYQLAPSE---ERYVMGKANSVFEDLSVTLR-QLNRRLFOENSV 164
Db 1779 LCTCSOLHSEETKQVLRSLTFESLPRKYLLFGISVKVSAVAIAFRSSYRDRSPSGSY 1838
QY 165 -LSSLPPLNLSNNNEVDLFLSELQVLDHDISSL-LSRHKHLAK-----DHSPLYSLELA 217
Db 1839 ERETFALTSLEVTYEA-LLEIME-ACMRDIPCKMLQWTELAQRARQYNPSLQPRALV 1896
QY 218 GLDEICKR--YGEDEQFRDASKIYDALQKPADMS-----253
Db 1897 VFGCISKRVSQIKQIIRILSKALESCLK--GPDYNSQVLIETATVIALTKLOPLLNKD 1954
QY 254 -----LYGNAVVEL--VTYKSPDTSIRKTRTILEAKQAKNPASP-----292
Db 1955 SPLKRALFWAVAVAYQLQDEVNLSAGTALLEQNLHTLDSLRIFNOKSPREVFMAIRNPLE 2014
QY 293 -----YNLAYKYNFEYSVFNMY 310
Db 2015 WHCKQMDHFVGLNFNSNFPALVGHLL 2041

```

```

RESULT 4
PCT-US94-00198-2
Sequence 2, Application PC/TUS9400198
GENERAL INFORMATION:
APPLICANT: Schering Corp.
TITLE OF INVENTION: RAS Associated GAP Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering Corp.
STREET: 1 Giraldo Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 6.0.8
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00198
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/004,824
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lunn, Paul G.
REGISTRATION NUMBER: 32,743
REFERENCE/DOCKET NUMBER: DX0352 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)822-7255
TELEFAX: (201)822-7039
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2485 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PCT-US94-00198-2

```

Query Match 5.7%; Score 100.5; DB 5; Length 2485;
 Best Local Similarity 22.2%; Pred. No. 1.1;
 Matches 86; Conservative 60; Mismatches 140; Indels 101; Gaps 18;

```

QY 13 CVCNFEETLAKSPGSVFRNGNMPRGE-----RIPDYAALSMGFSYKEDLSMEGLAV 65
Db 1667 GGLSITKAEVMAADPAVALASGNVKSVMGKCKIIDTCTSLPTPEQHLMDWDIAL 1726
QY 66 GNLFHPRATVAVWVKGVNKLALPG-----SVISYPLENAVFPISDS-----VANSHHS 115
Db 1727 -----LARYMLMSTFNNSLDVAHLPLYLHVYTF-LVATGPLSLRSTGVLINITHS 1778
QY 116 L-----FSEETPVYQLAPSE---ERYVMGKANSVFEDLSVTLR-QLNRRLFOENSV 164
Db 1779 LCTCSOLHSEETKQVLRSLTFESLPRKYLLFGISVKVSAVAIAFRSSYRDRSPSGSY 1838
QY 165 -LSSLPPLNLSNNNEVDLFLSELQVLDHDISSL-LSRHKHLAK-----DHSPLYSLELA 217
Db 1839 ERETFALTSLEVTYEA-LLEIME-ACMRDIPCKMLQWTELAQRARQYNPSLQPRALV 1896
QY 218 GLDEICKR--YGEDEQFRDASKIYDALQKPADMS-----253
Db 1897 VFGCISKRVSQIKQIIRILSKALESCLK--GPDYNSQVLIETATVIALTKLOPLLNKD 1954
QY 254 -----LYGNAVVEL--VTYKSPDTSIRKTRTILEAKQAKNPASP-----292
Db 1955 SPLKRALFWAVAVAYQLQDEVNLSAGTALLEQNLHTLDSLRIFNOKSPREVFMAIRNPLE 2014
QY 293 -----YNLAYKYNFEYSVFNMY 310
Db 2015 WHCKQMDHFVGLNFNSNFPALVGHLL 2041

```

```

RESULT 5
US-08-510-284-1
Sequence 1, Application US/08510284
Patent No. 5580955
GENERAL INFORMATION:
APPLICANT: Nur-E-Kamal, M. S. A.; Maruta, Hiroshi
TITLE OF INVENTION: FRAGMENTS OF NEUROFIBROMIN (NF1) AND METHOD
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,284
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/071,575
FILING DATE: 1-JUNE-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5580955man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5319
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 688-9884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2818 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PUBLICATION INFORMATION:
AUTHORS: Marchuk, Douglas A.; Saulino, Ann M.;

```

AUTHORS: Tavakkol, Roxanne; Swaroop, Manju;
 AUTHORS: Wallace, Margaret R.; Andersen, Lone B.;
 AUTHORS: Mitchell, Anna L.; Gutmann, David H.;
 AUTHORS: Boguski, Mark; Collins, Francis S.
 TITLE: CDNA Cloning of the Type 1 Neurofibromatosis Gene:
 TITLE: Complete Sequence of the NF1 Gene Product
 Patent No. 5580955
 JOURNAL: Genomics
 VOLUME: 11
 PAGES: 931-940
 DATE: 1991
 US-08-510-284-1

Query Match 5.7%; Score 100.5; DB 1; Length 2818;
 Best Local Similarity 22.2%; Pred. No. 1.3;
 Matches 86; Conservative 60; Mismatches 140; Indels 101; Gaps 18;

QY 13 GVLGNEFSILKSPGVSFVRNGNMPJPG-----RIPDVAALSMGFSVKEDLSMPGLAV 65
 DB 2000 GGLGSIKAEVADVAVALASGNVKKLVSSKVGIRCKTIIDKTCISPTPLEQHLMMDIAI 2059
 QY 66 GNLPHRRATVWVWVKVKNLALPPG-----SVISYLENAVPSLDS-----VANSIHS 115
 DB 2060 -----LARYMLMSFNNSLDVAHLPLYLHVYTF-LVATGPLSLRASTHGLVINITIHS 2111
 QY 116 L-----FSEETPVYQLAPSE---ERYVWVGKANSFEDLSVTLR-QLNRRLFOENSV 164
 DB 2112 LCTCQSLHSEETKQVLRSLTEFSLPKFYLLFGISKVSAAVIAFRSSYRDRSFSGSY 2171
 QY 165 -LSLPLNLSLNNEVDLLFSLLOYLDISSL--LSRKHKLAK-----DHSPLDISLELA 217
 DB 2172 ERETFALTSLETVEA-LLEIME-ACMRDIPYCKWLDQWTELAQRAPQYNSLQPRALV 2229
 QY 218 GLDEICKR--YGEDSEQFRDASKIIVDALOKFADWYS----- 253
 DB 2230 VFQICISKRVSHQIKQIIRILSKALESCKL--GPDYNSQVLEATVIALFKLQPLNMD 2287
 QY 254 -----LYGNAVEL--VTYKSPDTSILKRTITILEAKQAKNPASP----- 292
 DB 2288 SPLHKALEFWAVAYVQLDEVNLYSAGTALLEONLHTLDSLRIFNDKSPREYVMAIRNPLE 2347
 QY 293 -----YNLAYKYNFEYSVFNWV 310
 DB 2348 WHCKQMDHFVGLNFNSNFPALVGHLL 2374

RESULT 6
 US-08-411-389-2
 ; Sequence 2, Application US/08411389
 ; Patent No. 5605799
 ; GENERAL INFORMATION:
 ; APPLICANT: White, Raymond L.
 ; APPLICANT: Cawthon, Richard M.
 ; APPLICANT: Li, Ying
 ; TITLE OF INVENTION: SOMATIC MUTATIONS IN THE
 ; TITLE OF INVENTION: NEUROFIBROMATOSIS TYPE 1 GENE IN HUMAN TUMORS
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
 ; STREET: 1201 New York Avenue NW, Suite 1000
 ; CITY: Washington
 ; STATE: DC
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/411,389
 ; FILING DATE:
 ; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/047,088
 ; FILING DATE: 16-Apr-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Innen, Jeffrey L.
 ; REGISTRATION NUMBER: 28,957
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-962-4810
 ; TELEFAX: 202-962-8300
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2818 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-411-389-2

Query Match 5.7%; Score 100.5; DB 1; Length 2818;
 Best Local Similarity 22.2%; Pred. No. 1.3;
 Matches 86; Conservative 60; Mismatches 140; Indels 101; Gaps 18;

QY 13 GVLGNEFSILKSPGVSFVRNGNMPJPG-----RIPDVAALSMGFSVKEDLSMPGLAV 65
 DB 2000 GGLGSIKAEVADVAVALASGNVKKLVSSKVGIRCKTIIDKTCISPTPLEQHLMMDIAI 2059
 QY 66 GNLPHRRATVWVWVKVKNLALPPG-----SVISYLENAVPSLDS-----VANSIHS 115
 DB 2060 -----LARYMLMSFNNSLDVAHLPLYLHVYTF-LVATGPLSLRASTHGLVINITIHS 2111
 QY 116 L-----FSEETPVYQLAPSE---ERYVWVGKANSFEDLSVTLR-QLNRRLFOENSV 164
 DB 2112 LCTCQSLHSEETKQVLRSLTEFSLPKFYLLFGISKVSAAVIAFRSSYRDRSFSGSY 2171
 QY 165 -LSLPLNLSLNNEVDLLFSLLOYLDISSL--LSRKHKLAK-----DHSPLDISLELA 217
 DB 2172 ERETFALTSLETVEA-LLEIME-ACMRDIPYCKWLDQWTELAQRAPQYNSLQPRALV 2229
 QY 218 GLDEICKR--YGEDSEQFRDASKIIVDALOKFADWYS----- 253
 DB 2230 VFQICISKRVSHQIKQIIRILSKALESCKL--GPDYNSQVLEATVIALFKLQPLNMD 2287
 QY 254 -----LYGNAVEL--VTYKSPDTSILKRTITILEAKQAKNPASP----- 292
 DB 2288 SPLHKALEFWAVAYVQLDEVNLYSAGTALLEONLHTLDSLRIFNDKSPREYVMAIRNPLE 2347
 QY 293 -----YNLAYKYNFEYSVFNWV 310
 DB 2348 WHCKQMDHFVGLNFNSNFPALVGHLL 2374

RESULT 7
 US-08-449-933-2
 ; Sequence 2, Application US/08449933
 ; Patent No. 5859195
 ; GENERAL INFORMATION:
 ; APPLICANT: Collins, Francis S.
 ; APPLICANT: Wallace, Margaret R.
 ; APPLICANT: Marchuk, Douglas A.
 ; APPLICANT: Anderson, Lone B.
 ; APPLICANT: Guttman, David H.
 ; TITLE OF INVENTION: Neurofibromatosis Gene
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Morrison & Foerster
 ; STREET: 755 Page Mill Road
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304-1018
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

Db 2288 SPLHKALEFVAVAYQDDEVNLYSACTALLEONLHTLDSIRIFNDKSPPEYFMAIRNPLE 2347
 Oy 293 -----YNLAKYKNPEXSVFNNV 310
 Db 2348 WHCKQMDHFVGLNINSFNALVGHLL 2374

RESULT 8

US-07-966-049A-2
 Sequence 2, Application US/07966049A
 Patent No. 6238861
 GENERAL INFORMATION:
 APPLICANT: Collins, Francis S.
 APPLICANT: Wallace, Margaret R.
 APPLICANT: Marchuk, Douglas A.
 APPLICANT: Anderson, David H.
 APPLICANT: Gultman, David H.
 TITLE OF INVENTION: Neurofibromatosis Gene
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Morrison & Foerster
 STREET: 755 Page Mill Road
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304-1018
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/966,049A
 FILING DATE: 02-AUG-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Koski, Antoinette F.
 REGISTRATION NUMBER: 34,202
 REFERENCE/DOCKET NUMBER: 20344-20553.20
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 813-5600
 TELEFAX: (415) 494-0792
 TELEX: 706141 MRSNFDRS SFO
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2818 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHEetical: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT: 17q11.2
 FEATURE:
 NAME/KEY: Cleavage-site
 LOCATION: group(583..586, 815..818, 2573..2576, 2810..2813)
 OTHER INFORMATION: /note= "Potential CAMP-dependent
 OTHER INFORMATION: protein kinase recognition sites"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 2549..2556
 OTHER INFORMATION: /note= "Potential tyrosine
 OTHER INFORMATION: phosphorylation site"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: group(1264, 1276, 1358, 1377, 1389, 1390, 1391,
 LOCATION: 1395, 1396, 1400, 1423, 1426, 1429, 1430)
 OTHER INFORMATION: /note= "Invariant residues within
 OTHER INFORMATION: most statistically significant regions of similarity among th

OTHER INFORMATION: GAP family of proteins"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: group(1264..1290, 1345..1407, 1415..1430)
 OTHER INFORMATION: /note= "Most statistically
 OTHER INFORMATION: significant regions of similarity among the GAP family of
 OTHER INFORMATION: proteins"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 496
 OTHER INFORMATION: /note= "At variance with previously
 OTHER INFORMATION: published sequence which shows an ATG methionine codon rat
 OTHER INFORMATION: than an ATA isoleucine codon"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 1183
 OTHER INFORMATION: /note= "At variance with previously
 OTHER INFORMATION: published sequence. Shows an CTG leucine codon rather than
 OTHER INFORMATION: previously published CTC"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 1555
 OTHER INFORMATION: /note= "At variance with previously
 OTHER INFORMATION: published sequence. Lacks an extra CAT histidine condon at
 OTHER INFORMATION: this residue"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: (2771-2772)
 OTHER INFORMATION: /note= "Position of an 18 amino
 OTHER INFORMATION: acid insertion(SEQ ID NO:10) representing an alternatively
 OTHER INFORMATION: spliced product"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: (1376-1377)
 OTHER INFORMATION: /note= "Position of a 21 amino acid
 OTHER INFORMATION: insertion representing an alternatively spliced product"
 FEATURE:
 NAME/KEY: Domain
 LOCATION: 1125..1537
 OTHER INFORMATION: /note= "NF1 catalytic domain"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 2746..2818
 OTHER INFORMATION: /note= "Corresponding amino acids
 OTHER INFORMATION: for the PstI-HindIII fragment designated pMAL.B3A"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 65..371
 OTHER INFORMATION: /note= "Corresponding amino acids
 OTHER INFORMATION: for the HpaI-PstI fragment designated pMAL.HF3A.p"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 65..1240
 OTHER INFORMATION: /note= "Corresponding amino acids
 OTHER INFORMATION: for the HpaI-XhoI fragment designated pMAL.HF3A.X"
 PUBLICATION INFORMATION:
 AUTHORS: Wallace, M.R. et al.
 TITLE: Type I Neurofibromatosis Gene: Correction
 JOURNAL: Science
 VOLUME: 250
 ISSUE: 12/21/90
 PAGES: 1749-
 DATE: 12/21-1990
 RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2818
 PUBLICATION INFORMATION:
 AUTHORS: Wallace, M.R. et al.
 TITLE: Type I Neurofibromatosis Gene: Identification
 JOURNAL: Science
 VOLUME: 249
 ISSUE: 07/13/90
 PAGES: 181-186
 DATE: 07/13-1990

Db 1226 LKEFGALPSTYRSLWELQQRLPKSA-----EST-----HPGQTVLI 1264

```
QY 80 VKGVAKLALPGOSVISTYPLENAV-----FSLDSVANSIHSLTSEETPVVILQLABEER 133
```

DB 1265 IDGADRLVDQNGQLISDNIIPKLLRCVHLVLSVSSDAGIGETLEDSQGAHVLALGLEAS 1324

```

02      VIMQKANSVEEDLSVLNQLKNRKHQENSVL-----SSLPLNSLRINNEVDLPFSE- 186
134      :|::||:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1325     -AABDIIDDEIATWQDTFFHQNNNOUHTTIRWEQQOQDIXETIEUEEETITRWHS 3000

```

187 -----LOWI HODGETT SEDH--KITT ANDUSODD VST BU KOT DITOC DRYKATOC OMOHNA--OWTT 040
DU AKANLVKRELLALIOARKLEESPEFNQOMRLELLVKRKSGRPYLDKLVIDHARKLETFVEQ 1386

[illegible]

241 VDA10KFADDDMYSLYGGNAVVETLVYKSPDPTLIRKTRTTILFAKQAKNPASPY 293

Db 1424 -----SGLTVDQLHGVLVSVRTLPKGTIKSMEAEVAAGNSGDPY 1461

US-09-143-954-4
RESULT 12

Sequence 4, Application US/09143954A
Patent No. 6194170

```

; GENERAL INFORMATION:
;
; APPLICANT: Wallis, Nicola G.
;
; DATED OF INVENTION: 2007

```

FILE REFERENCE: GM10097
CURRENT APPLICATION NUMBER: JIS/09/143 954A

EARLIER APPLICATION NUMBER: 60/060-011
 CURRENT FILING DATE: 1998-08-31
 INVENTOR: CO/03/143, 536A
 CLASSIFICATION:

```

: EARLIER FILING DATE: 1997-09-25
:
: NUMBER OF SEQ ID NOS: 6

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4

```

```

; LENGTH: 448
; TYPE: PRt

```

ORGANISM: *Streptococcus pneumoniae*
US-09-143-954-4

Query match	5.2%;	Score 91;	DB 4;	Length 448;
Best Local Similarity	22.5%;	Pred. No. 0.61;		
Matches	48	Correctly	36	55

98	LENAPDESIVANSTIISTEETNYIT OI ADEEPNIVATVAVGUREDT QUTET FOR DNE	157
molecules	40; conservative	39; mismatches
		60; indels
		60; caps

252 LEQALDLYEYTGKYNATNMTA--SYVALIOEGVSEFQTH-----QAEQCTET-----PND 300

Yr	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
Yr	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
Yr	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
Yr	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175																																																																																																																													

```

Db          : |          : |          : |          : |          : |
300 TEWKAANGADILSDVYNANPTAMKLIETFSALPANEGCK---JAVLADMKEIGDOS 355

```

195 SLLSRHKLAKDHPDL-----YSLELAGLDE-----IGKRY-----GEDSEOFRDAS 237

Db 356 -- IQLHNQMLSLSPDVLDTYIFYGEDLAEIQAQLSQMFPGHVYFEKTEDDQDFEDLV 413

QY 238 KILVDAIÖKFADDMSLYGNAVELYVVKSF 270

Db 414 KÖVESÉS--ANDÖILLKGSNSMNLAMLVESLE 444

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

US-09-143-954-2
Sequence 2. Annotation ITS/09143954A

GENERAL INFORMATION:
Patent No. 6194170

APPLICANT: wallis, Nicola G.
TITLE OF INVENTION: MURF

FILE REFERENCE: GM10097
CURRENT APPLICATION NUMBER: US/09/143,954A

```

; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 60/060,011

```

EARLIER FILING DATE: 1997-09-25

[illegible]

```

: NUMBER OF SEO ID NOS: 6
: SOFTWARE: FastSEO for Windows Version 4.0.
: SEO ID NO 2
: LENGTH: 457
: TYPE: PRT
: ORGANISM: Streptococcus pneumoniae
US-09-143-954-2

```

Query Match 5.28; Score 91; DB 4; Length 457;
Best Local Similarity 22.5%; Pred. No. 0.63;
Matches 48; Conservative 39; Mismatches 66; Indels 60; Gaps 11.

```

OY 1ENAPESLSVANSJHSJFSETPVYLQALPEEYVYVAKANSVEDJSLYLQJLNR 157
    | | : : : : : | | | | | : : : : :
Db 261 LEQALDPEYTKKNAINAMIA--SYVALQSGVSEEDIH-----QAFQGIELT-----RNR 308
                                         -----LFOEN-----
OY 158 -----SYLSLSPUNLSIRNNEVDLLPLSELQVHDIS 194
Db 309 TEWKKAANGADILSDYYNANPFAKMLILETFSSIAINEGCK-----IAVLADWKELSDQS 364
                                         : : : : :
OY 195 SLSNHHKHLAKHSPL-----YSLELALGDE-----IGKRY-----GESEQFRRAS 237
    | | : : : : : | | : : : : : | | : : : : :
Db 365 --TQAHNMILSLSPVDLYLFYEGEDIAELQALASOMFPIGHVYFKKTEQDQDFEDLV 422
                                         : : : : :
OY 238 KIVLDALQKFAADWMSLYGCAVAVELTVYSPD 270
    | : : : : | : : | : : : : | : : : :
Db 423 KQYKESLS--ANDQILKSGSNMILAWYSESL 453

```

RESULT 14
US-08-952-127-12
; Sequence 12, Application US/08952127

GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
APPLICANT: Tagle, Danilo A.
APPLICANT: Collins, Francis S.
TITLE OF INVENTION: ATX1A-TELANGLECTASIA GENE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: Kohn & Associates
STREET: 30500 No. 6211361western Hwy., Suite 4100
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/08/952,127

CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REFERENCE NUMBER: 30,995
TELECOMMUNICATION NUMBER: 2290,000230
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5055
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3066 amino acids
TYPE: amino acid
SPRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Mus musculus
OS-08-952-127-12

Query Match	5.28;	Score 91;	DB 4;	Length 3066;
Best Local Similarity	23.1%;	Pred. No. 15;		
Matches 50;	Conservative 41;	Mismatches 97;	Indels 28;	Gaps 7.

```

QY 139 KANVEED,SVTLROLRNR,LOENSVALSSI,PLANSI,SNRNVLDLFL,SELOVJLHSSLS 198
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2126 ECTSHSLSYLNALQCCRRNFE--STFESYLRASLFRKKEVEBELSGSLESVSLYPTTS 2188
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 199 RHKHLAK-DHSPLYLSLALDELICGRYGEDSEQFRDASKLIVDALOKFADDMYSILYGG 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2184 RLQALIGLEHNSGSLTSRSVT-----DRENSLAWKWKQKHQLKDSDFSPOEPMLAL--R 2233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 256 NAYVELVYKSEDTSLIRKRTTILE-----AKQAKNPASPYNLAYK---YN----- 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2237 TIVLETLVOKEMERGOSAGCSKDILTKHLVEFSVLARTFKKNTDPERAIFKIKOYNSAIG 2296
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 301 -PEYSVVENVYLMALALAVITTSNTNMNDPXY 335
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2297 ISEMHLEAGVFNAKKESDLSLSTLQMKIRKLDDSEF 2332
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 15
US-09-063-950-2
: Sequence 2, Application US/09063950C
: Patent No. 6225085
: GENERAL INFORMATION:
: APPLICANT: Holtzman, Douglas A.
: TITLE OF INVENTION: NOVEL LRSG PROTEIN AND NUCLEIC ACID MOLECULES AND USES
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: MEI-019
: CURRENT APPLICATION NUMBER: US/09/063,950C
: CURRENT FILING DATE: 1998-04-21
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 673
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-063-950-2

```

Query Match	5.18;	Score 90.5;	DB 4;	Length 673;
Best Local Similarity	24.68;	Pred. No. 1.4;		
Matches 75; Conservative	39;	Mismatches 108;	Indels 83;	Gaps 16;

[illegible]

Mon Aug 20 10:21:55 2001

us-09-284-320-6.ra1

Page 10

Search completed: July 3, 2001, 16:18:53
Job time: 165 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 18, 2001, 18:14:09 ; Search time 5436.15 Seconds
(without alignments)
2987.617 Million cell updates/sec

Title: US-09-284-320-31

Sequence: 1 atgagctgtgttctgtctgtct.....accagaagatctgaatgat 1050

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_p11:*
13: gb_p12:*
14: gb_p13:*
15: gb_p14:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_hugo_hum:*
20: em_hugo_inv:*
21: em_hugo_rod:*
22: em_hug_hum1:*
23: em_hug_hum2:*
24: em_hug_hum3:*
25: em_hug_hum4:*
26: em_hug_hum5:*
27: em_hug_hum6:*
28: em_hug_hum7:*
29: em_hug_hum8:*
30: em_hug_inv1:*
31: em_hug_inv2:*
32: em_hug_other:*
33: em_hug_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_p1:*
48: em_rod:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_hug1:*
61: gb_hug2:*
62: gb_hug3:*
63: gb_hug4:*
64: gb_hug5:*
65: gb_hug6:*
66: gb_hug7:*
67: gb_hug8:*
68: gb_hug9:*
69: gb_hug10:*
70: gb_hug11:*
71: gb_hug12:*
72: gb_hug13:*
73: gb_hug14:*
74: gb_hug15:*
75: gb_hug16:*
76: gb_hug17:*
77: gb_hug18:*
78: gb_hug19:*
79: gb_hug20:*
80: gb_hug21:*
81: gb_hug22:*
82: gb_hug23:*
83: gb_hug24:*
84: gb_hug25:*
85: gb_p11:*
86: gb_p12:*
87: gb_p13:*
88: gb_p14:*
89: gb_p15:*
90: gb_p16:*
91: gb_p17:*
92: gb_p18:*
93: gb_p19:*
94: gb_rod:*
95: gb_rod2:*
96: gb_in4:*
97: gb_p10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1048.4	99.8	2049	89	AF248966
2	990.4	94.3	1884	93	BSM80272
3	860.4	81.9	158239	80	AL356315
4	860.4	81.9	214692	81	AL513325
5	858.8	81.8	165617	80	AL356381
6	491	46.8	622	97	HSY17975
7	193	18.4	126543	69	AC026156
8	152	14.5	126543	69	AC026156

C	9	132.2	12.6	77831	94	AF259074	AF259074	Mus musc
C	10	57	5.4	162337	81	AL512594	AL512594	Mus musc
C	11	56.2	5.4	76748	63	AC011256	AC014236	Drosophila
C	12	56.2	5.4	146847	4	AC008234	AC008234	Drosophila
C	13	56.2	5.4	162595	4	AC008356	AC008356	Drosophila
C	14	50.8	5.4	224400	5	AE003682	AE003682	Drosophila
C	15	56.2	4.8	158537	93	HSAC020696	AC002069	Human BAC
C	16	41.6	4.0	64743	72	AC043386	AC043836	Homo sapi
C	17	41.4	3.9	155003	78	AC090669	AC090669	Homo sapi
C	18	41.4	3.9	157400	64	AC016104	AC016104	Homo sapi
C	19	41.4	3.9	193007	66	AC021195	AC021195	Homo sapi
C	20	40.2	3.8	98543	85	AC005352	AC005352	Homo sapi
C	21	40.2	3.8	126732	62	AC001138	AC001138	Homo sapi
C	22	40	3.8	152713	69	AC026132	AC026132	Homo sapi
C	23	39.8	3.8	190125	78	AL138689	AC018689	Homo sapi
C	24	39	3.7	102064	85	AC002558	AC002558	Homo sapi
C	25	39	3.7	157135	72	AC064817	AC064817	Homo sapi
C	26	38.8	3.7	10463	12	AF0008125	AF0008125	Arabidops
C	27	38.8	3.7	82281	12	AC025295	AC025295	Arabidops
C	28	38.4	3.7	155992	71	AC034184	AC034184	Homo sapi
C	29	38.4	3.7	155992	76	AC083854	AC083854	Homo sapi
C	30	38.4	3.7	171864	65	AC018919	AC018919	Homo sapi
C	31	38.4	3.7	196758	61	AC009653	AC009653	Homo sapi
C	32	38	3.6	7218	10	166494	166494	Sequence
C	33	37.8	3.6	175483	65	AC018553	AC018553	Homo sapi
C	34	37.8	3.6	196775	72	AC064793	AC064793	Mus musc
C	35	37.6	3.6	8021	10	AX083744	AX083744	Sequence
C	36	37.6	3.6	80217	63	AC011840	AC011840	Drosophila
C	37	37.6	3.6	174894	4	AC007984	AC007984	Drosophila
C	38	37.6	3.6	179486	4	AC008216	AC008216	Drosophila
C	39	37.6	3.6	226471	5	AE003759	AE003759	Drosophila
C	40	37.4	3.6	593	13	AF202647	AF202647	Avena sativ
C	41	37.2	3.5	175144	73	AC068721	AC068721	Homo sapi
C	42	37.2	3.5	179852	79	AL353780	AL353780	Homo sapi
C	43	37	3.5	118007	60	AC008410	AC008410	Homo sapi
C	44	37	3.5	175156	74	AC069435	AC069435	Homo sapi
C	45	36.8	3.5	2184	12	AF123852	AF123852	Gelium gn

ALIGNMENTS

AF09074 Mus musc
AL51254 Mus musc
AC014236 Drosophila
AC000823 Drosophila
AC000836 Drosophila
AF003682 Drosophila
AC002069 Human BAC
AC048368 Homo sapi
AC090669 Homo sapi
AC015104 Homo sapi
AC021195 Homo sapi
AC005352 Homo sapi
AC001138 Homo sapi
AC026132 Homo sapi
AL183668 Homo sapi
AC002558 Homo sapi
AC064817 Homo sapi
AF008125 Arabidops
AC025295 Arabidops
AC034184 Homo sapi
AC038354 Homo sapi
AC018919 Homo sapi
AC009653 Homo sapi
16694 Sequence 14
AC018553 Homo sapi
AC064793 Mus musc
AX083744 Sequence
AC014840 Drosophila
AC000798 Drosophila
AC000821 Drosophila
AF003759 Drosophila
AF026247 Avena sativa
AC068721 Homo sapi
AL353790 Homo sapi
AC008410 Homo sapi
AC069435 Homo sapi
AF123652 Gmelin gn

RESULT	1					
LOCUS	AF248966	2049 bp	mRNA			02-JAN-2001
DEFINITION	Homo sapiens HT028 mRNA, complete cds.					
ACCESSION	AF248966					
VERSION	AF248966.1	GI:12005668				
KEYWORDS	FLI_CDNA.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
TITLE	1 (bases 1 to 2049)					
JOURNAL	Yang, Y., Xu, X., Gao, G., Xiao, H., Chen, Z., and Han, Z.					
	Direct Submission					
	Submitted (27-MAR-2000) Chinese National Human Genome Center at					
	Shanghai 351 Guo Shoujing Road, Zhongjiang Hi-Tech Park, Pudong,					
	Shanghai 201203, People's Republic of China					
FEATURES	Location/Qualifiers					
source	1..2049					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/tissue_type="hypothalamus"					
	108..1160					
	/codon_start=1					
	/product="HT028"					
	/protein_id="AAG4564.1"					
	/db_xref="GI:12005669"					
	/translation="MAVFVLLALVAGLVGNEFLTSPGVYFRNGWPIPIGER					
	VALALSGPVEKEDLSFSGLVAGNLPHRPRTATWVQYCNKLALPGVSTYPLEH					
	PESLDVANGYHSLFSSETFPVYLALASERTYTMGANGVAFDLSVTLRQNRRI					
	ENSVLSPLNLSRNNEVDLFLSELYQVYHDISSLSRKLHADKSPDLYSELYSE					

[illegible]

Db 1008 TTGCAATATTCGCTGCTTTCACATGCTACTTTCATATGATCCCTTGCCCTTGCT 1067
Qy 961 ggtattatcactcttaacaatttggaaacatgagccctgagatagatagcaatctat 1020
Db 1068 GTGATTATACCTCTTTCACATATTTTGAACATGATGCTGATATGATACATCATTTAT 1127
Qy 1021 aggatgacaacacgaagattcgaaatgat 1050
Db 1128 AGGATGACAAACGACGATTCGATGAT 1157

RESULT 2

HSMB00272 1884 bp mRNA PRI 18-FEB-2000
LOCUS Homo sapiens mRNA: cDNA DKFZp564O0582 (from clone DKFZp564O0582);
DEFINITION partial cds.

ACCESSION AL049929.1 GI:4884174

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 1884)
Poustka A., Klein M., Mewes H.W., Gassenhuber J. and Wiemann S.
Submitted (15-MAY-1999) MIPS, Am Klopferplatz 18a, D-82152
Martinsried, GERMANY

COMMENT

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp564O0582) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES

SOURCE

1..1884
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="Fetal"
/tissue_type="Brain"
/clone="DKFZp564O0582"
/clone_1kb="564 (synonym: hfbp2). Vector pAMP1; host
XL-2blue; sites NotI + SalI"
/map="x"
1..995
/gene="DKFZp564O0582"
<1..995
/note="weak similarity to C.elegans T24D5.2"
/note="weak similarity to C.elegans T24D5.2"
/product="hypothetical protein"
/protein_id="CAB43210.1"
/db_xref="GI:4884175"
/translation="TAKSPGVFRNMPFIEBRIPIYVALSKGSVKEKEDLSWFGLA
VGNLHPRATVWVWVGVKLVLPKSVSTPLENAVPSLDVANSVHSLFSEETP
VVLQASERVYVWVKANSVFDLSVTLKLNRLFOENSVLSLPLNSLNNEVD
LFLSELQVLDHISLSLRHKLAKHSDPVLSTLEAGLIDIGKRYGENDSEDFRASK
LLVDALQFADDMYSLYGNNAVELVYKSFDSLRTKRTILAKKAPKAPYNLA
KYNREXSVYENVMIMIALAVIITSNIMNDPVDISIIIRMNQKIRND"

gene

CDS

polya_signal
polya_site
BASE COUNT 555 a 322 c 393 g 614 t
ORIGIN

Query Match 94.3%; Score 990.4; DB 93; Length 1884;
Best Local Similarity 99.9%; Pred. No. 4,6e-281;
Matches 991; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 59 gtaataaataccacagagctgtgtgttttcgaaatggaaattggcctatccacagag 118
Db 1 GATATTAAATCAACAGAGGCTGTGTGTTTCCAAATVGAATGGCTTATACAGAG 60
Qy 119 agcgaatccacagacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac 178
Db 61 AGCGGATGCCAGAGGCTGAGTGTGTCATGAGGCTTCTGTGAAAGAGACCTTCTT 120
Qy 179 gcccaagactgcagtggttaacgtgttcaatgctcctgggtcagctcagtgatg 238
Db 121 GCCCAGAGCTGCGACGTGGGTAACTGTTCATGCTCCGCGCTACCGCTAGGTGTG 180
Qy 239 tgaaggagtgaaacactgctcctaccccaaggcagtgacatcctgacatccttgaga 298
Db 181 TGAAGGAGTGAACAACTGCTCTACCCCGAGGAGTGTATGATACCTTTGGAGA 240
Qy 299 atgcactcctcttaagcttgacagtggttgcaaatcctacatcctcctatcttga 358
Db 241 ATGCACTCTCTTTAGTCTTGACAGTGTGCAAAATCCATTCACCTTATTTCTGAG 300
Qy 359 aaactcgtgtgttgacgttggtcctcccaagtgaagagtgatataggtaaggag 418
Db 301 AAACCTCTGTTTGTGAGTGTGCTCCAGTGAAGAGATGATATGATGAGGAGG 360
Qy 419 caaacctggttggaagacacttgacacttgacacttgacacttgacacttgacact 478
Db 361 CAACTCAGTGTGTGAAGACCTTTCAGTACCTTGCCGACGCTCCTAATGCTGTTTC 420
Qy 479 aagaaacactgttctcagctcactccctcctcctcctcctcctcctcctcctcctc 538
Db 421 AAGAAACCTCTGTTGAGTGTGCTCCAGTGAAGAGATGATATGATGAGGAGG 480
Qy 539 acctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 598
Db 481 ACCGCTCTCTTCTTCTGAACTGCAAGTGTGCAATGATTTTCAACTGCTGCTGTC 540
Qy 599 ataagcatcagcgaagatcattcctcctcctcctcctcctcctcctcctcctcct 658
Db 541 ATAGCATCTACCAAGATCATCTCTGATTTATATTCATGAGTGTGAGGAGGATTTG 600
Qy 659 atgaatctggaagagtgatagtggaagactcgtgaacattcgaagtgcttctaagtc 718
Db 601 ATGAATTTGGAAGGCTTATGAGGAGGAGTGTGCAATTTCAAGTGTGCTTAAATTC 660
Qy 719 ttgtgacgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 778
Db 661 TTGTTGACGCTCTGCAAAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGT 720
Qy 779 tggtagagtgatgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac 838
Db 721 TGGTAGAGTGTGTCTCACTGTCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTG 780
Qy 839 tccctgagcgaacaaagcgaagacacacacacacacacacacacacacacacacac 898
Db 781 TCTTGTAGGCAAAACGAGCAGGAGACCCAGCAAGTCTTAACTTGCATTAAGATATA 840
Qy 899 atttgaataatcgttggttttaacatgtaacttgtaataagatgacgtcgttgct 958
Db 841 ATTTGAATATTCCTGCTTTTCAACATGCTGATTTGATTAAGATGATGCTGCTGG 900
Qy 959 ctggtgatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1018
Db 901 CTGTGATTTACACCTCTTACATATTGGACATGATGATGATGATGATGATGATGAT 960
Qy 1019 atagatgacaaacacacacacacacacacacacacacacacacacacacacac 1050
Db 961 ATAGATGACAAACCAAGATTCGATGATGAT 992

RESULT 3
AL356315 158239 bp DNA HTG 22-MAR-2001
LOCUS Homo sapiens chromosome 1 clone RP11-332F5, *** SEQUENCING IN
DEFINITION

```

ACCESSION      AL356315      6 unordered pieces.
VERSION        AL356315.8   GI:13445377
KEYWORDS       HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 158239)
AUTHORS        Plumb,B.
TITLE          Direct Submission
JOURNAL        Submitted (21-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
                requests: clonerequests@sanger.ac.uk
                On Mar 24, 2001 this sequence version replaced gi:13443378.
COMMENT        ----- Genome Center
                Center: Sanger Centre
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquerry@sanger.ac.uk
                ----- Project Information
                Center project name: ba332f5
                ----- Summary Statistics
                Assembly program: XGAP; version 4.5
                Sequencing vector: plasmid; L08752; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Consensus quality: 156356 bases at least Q40
                Consensus quality: 156868 bases at least Q30
                Consensus quality: 157303 bases at least Q20
                Insert size: 157739; sum-of-contigs
                Insert size: 143450; 18.0% error; agarose-fp
                Quality coverage: 6.86x in Q20 bases; sum-of-contigs Quality
                coverage: 7.61x in Q20 bases; agarose-fp
                -----
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 6 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                *
                1 22704: contig of 22704 bp in length
                * 22705 22804: gap of 100 bp
                * 22805 26580: contig of 3776 bp in length
                * 26581 26680: gap of 100 bp
                * 26681 53453: contig of 26775 bp in length
                * 53456 53553: gap of 100 bp
                * 53556 58106: contig of 4551 bp in length
                * 58107 58206: gap of 100 bp
                * 58207 78538: contig of 20332 bp in length
                * 78539 78638: gap of 100 bp
                * 78639 158239: contig of 79601 bp in length.
FEATURES
source
    1..158239
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /chromosome="1"
     /clone_id="RP11-332F5"
     /clone_1fp="RPC1-11.2"
     1..22704
      /note="assembly_fragment:00466
      fragment_chain:1
      clone_end:SP6
      vector_side:left"
      22805..26580
       /note="assembly_fragment:00415
       fragment_chain:1"
       26681..53455
        /note="assembly_fragment:01627
        fragment_chain:1"
        53556..58106
         /note="assembly_fragment:01215
         fragment_chain:1"

```


REFERENCE 2 (bases 1 to 622)
 AUTHORS Ludwig, J.H.
 TITLE Direct Submission
 JOURNAL Submitted (20-AUG-1998) J.H. Ludwig, Universitaetsklinikum
 Frankfurt, Gustav Emden-Zentrum der Biologischen Chemie (ZBC),
 Institut fuer Biochemie I, Theodor Stern-Kal 7 Haus 25 B, 60590
 Frankfurt, FRG
 COMMENT Related entry: R78826.
 FEATURES
 source Location/Qualifiers
 1..622
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 192..494
 /codon_start=1
 /product="putative vacuolar proton ATPase membrane sector
 associated protein M8-9"
 /protein_id="CAA76984.1"
 /db_xref="GI:3451261"
 /db_xref="SPTREMBL:O75787"
 /translation="MYSLGNNVVELVTVKSFDTSLIRKRTILEAKAKNPSPYN
 LAKYNEISVVFNNWLMIALALAVIITSYINIMNDPGYDSIIYRMTNOKIRMD"
 BASE COUNT 186 a 103 c 134 g 199 t
 ORIGIN
 Query Match 46.8%; Score 491; DB 97; Length 622;
 Best Local Similarity 100.0%; Pred. No. 7.9e-134; Mismatches 0; Indels 0; Gaps 0;
 Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 560 tgcgaagtctatcatgataatttcaagctgtgtctgtctcaagaatcagcaaggatc 619
 Db 1 TCCAACTGCTCAAGATATTTTCAAGCTGTCTGCTCTCAATAGCATCTACCAAGATC 60
 QY 620 attcctcattatattacacagctgagcaggtttgagtaaatgttggaagcgttaag 679
 Db 61 ATTCTCGATTATTTATTCACCTGACCTGACAGCTTTGAGTGAATTTGGAGAGGTTATG 120
 QY 680 ggggaagactcgaacaattcaagatgcttcttaagatccttltgaagctctgcaaat 739
 Db 121 GGGAAAGACTGGAACAATTCAGAGATGCTTAAAGATCTTGTGACCTCTGCAAAAGT 180
 QY 740 ttgcagatgacatgacagctcttattgttggaatgagtgatgagtgatgacgtgtca 799
 Db 181 TTGCAAGATGACATGTACAGCTTTATGGTGGGATGCAAGTGAAGTGAATCACTGCTCA 240
 QY 800 agtcatcttggaacctccctcattatggaagacaagactatccttgaaggcaaaacagca 859
 Db 241 AGTCATTTGACACCTCCCTCATTTAGGAAGACACGACTATCTTGAGCAAAACAGCA 300
 QY 860 agaacccagcaagctccctataacacctgacataaagtaatttgaatattcgtgtgt 919
 Db 301 AGAACCCAGCAAGTCCCTATAACCTTGATATAGTATTAATTTGAATATTCCTGCTGTT 360
 QY 920 tcaaatgttactcttgataatgataatgccttgagccttgagctgtgattatcaacttaca 979
 Db 361 TCAACATGTGACTTTGATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTACA 420
 QY 980 atatttggaaatgataatccttgataatgataatgataatgataatgataatgataatgata 1039
 Db 421 ATATTGGAAATGATGCTGCTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 QY 1040 ttcgaatgat 1050
 Db 481 TTGGAATGAT 491
 RESULT 7
 LOCUS AC026156 126543 bp DNA HTG 28-MAR-2000
 DEFINITION Homo sapiens chromosome 3p clone RP11-169016, WORKING DRAFT
 SEQUENCE, 10 unordered pieces.
 ACCESSION AC026156
 VERSION AC026156.2 GI:7331308

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 126543)
 AUTHORS Lu, G., Zhao, Y., Lin, W., Dong, H., Wan, M., Xu, S., Gu, W., Tu, Y.,
 Jia, J., Wu, C., Zhang, C., Zhong, M., Zhou, Y., Ren, S., Fu, G., Chen, Z.,
 and Huang, M.
 COMMENT Direct Submission
 Submitted (21-MAR-2000) Genomic Dept., Chinese National Human
 Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai,
 Shanghai 201203, P. R. China
 On Mar 28, 2000 this sequence version replaced gi:7271973.
 -----Genome Center Information-----
 Center: Chinese National Human Genome Center at Shanghai
 Web site: http://www.chgc.sh.cn
 Email: miuange@chgc.sh.cn or fuguang@chgc.sh.cn
 -----End Genome Center Information-----
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 3411: contig of 3411 bp in length
 * 3412 7987: contig of 4576 bp in length
 * 7988 16490: contig of 8503 bp in length
 * 16491 30218: contig of 13728 bp in length
 * 30219 44895: contig of 14677 bp in length
 * 44896 55628: contig of 10733 bp in length
 * 55629 71833: contig of 16205 bp in length
 * 71834 84537: contig of 12704 bp in length
 * 84538 103706: contig of 19169 bp in length
 * 103707 126543: contig of 22837 bp in length.
 FEATURES
 source 1..126543
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3p"
 /clone="RP11-169016"
 BASE COUNT 34774 a 28466 c 28425 g 34834 t 44 others
 ORIGIN
 Query Match 18.4%; Score 193; DB 69; Length 126543;
 Best Local Similarity 100.0%; Pred. No. 2.2e-45; Mismatches 0; Indels 0; Gaps 0;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 858 gaagaaccagcaagctccctataacacctgacataaagtaatttgaatattcgtgtgt 917
 Db 81178 GAAGAACCAGCAAGTCCCTATAACCTTGATATAGTATTAATTTGAATATTCCTGCTGTT 81237
 QY 918 ttcaaatgttactcttgataatgataatgccttgagccttgagctgtgattatcaacttaca 977
 Db 81238 TTTCAACATGTGACTTTGATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTACA 81297
 QY 978 caatatttggaaatgataatccttgataatgataatgataatgataatgataatgataatgata 1037
 Db 81298 CAATATTGGAAATGATGCTGCTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 81357


```
repeat_region      13291. .13464  
/rpt_family="TGC)n"  
repeat_region      14374. .15028  
/rpt_family="RLTR13C"  
repeat_region      14515. .15028  
/rpt_family="RLTR13D"  
repeat_region      15157. .15862  
/rpt_family="L1_MM"  
repeat_region      15873. .16356  
/rpt_family="L1F_Send"  
repeat_region      16085. .16568  
/rpt_family="L1F_Send"  
repeat_region      16509. .16992  
/rpt_family="L1F_Send"  
repeat_region      16933. .17341  
/rpt_family="L1F_Send"  
repeat_region      17449. .23342  
/rpt_family="L1_MM"  
repeat_region      23348. .23928  
/rpt_family="L1_MM"  
repeat_region      23365. .24582  
/rpt_family="L1_MM"  
repeat_region      24364. .24701  
/rpt_family="L1_MM"  
repeat_region      24599. .26220  
/rpt_family="L1_MM"  
repeat_region      complement(26217. .27218)  
/rpt_family="L1_MM"  
repeat_region      complement(27118. .27355)  
/rpt_family="L1F_Send"  
repeat_region      27881. .28092  
/rpt_family="B3"  
repeat_region      complement(28543. .28587)  
/rpt_family="(CA)n"  
repeat_region      complement(28929. .28969)  
/rpt_family="(CA)n"  
repeat_region      28938. .28969  
/rpt_unit="(GT)x15  
repeat_region      29384. .29531  
/rpt_family="MTD"  
repeat_region      29556. .29866  
/rpt_family="L1_MM"  
repeat_region      29875. .30280  
/rpt_family="RLTR13C"  
repeat_region      30282. .31643  
/rpt_family="L1_MM"  
repeat_region      31674. .32181  
/rpt_family="RLTR13C"  
repeat_region      32186. .34760  
/rpt_family="L1_MM"  
repeat_region      34747. .36147  
/rpt_family="L1_MM"  
repeat_region      34929. .34953  
/rpt_unit="(AAC)x6  
repeat_region      34930. .34953  
/rpt_family="(CAA)n"  
repeat_region      36156. .36306  
/rpt_family="Lx3"  
repeat_region      36337. .36557  
/rpt_family="RMER13A"  
repeat_region      36561. .37025  
/rpt_family="RMER13A"  
repeat_region      37029. .37841  
/rpt_family="Lx4"  
repeat_region      37821. .38450  
/rpt_family="Lx4"  
repeat_region      complement(38431. .38476)  
/rpt_family="AT_rich"  
repeat_region      39113. .39169  
/rpt_unit="(CA)x28  
repeat_region      39114. .39169  
/rpt_family="(CA)n"  
repeat_region      complement(39309. .40536)
```

```
repeat_region      /rpt_family="L1_MM"  
complement(40537. .40629)  
/rpt_family="Lx5"  
repeat_region      40636. .40687  
/rpt_family="RMER6A"  
repeat_region      40688. .41215  
/rpt_family="RMER6A"  
repeat_region      complement(41325. .41441)  
/rpt_family="RMER1B"  
repeat_region      41452. .41573  
/rpt_family="ORR1A3"  
repeat_region      42854. .42984  
/rpt_family="B1-F"  
repeat_region      42986. .43037  
/rpt_family="(CAA)n"  
repeat_region      complement(44550. .44693)  
/rpt_family="RMER1B"  
repeat_region      complement(44699. .44893)  
/rpt_family="RMER1B"  
repeat_region      44997. .45152  
/rpt_family="MERVL"  
repeat_region      45066. .45164  
/rpt_family="MERVL"  
repeat_region      45202. .46018  
/rpt_family="MERVL"  
repeat_region      complement(46029. .46760)  
/rpt_family="L1_MM"  
repeat_region      46762. .46965  
/rpt_family="MERVL"  
repeat_region      complement(46986. .47171)  
/rpt_family="ORR1A1"  
repeat_region      complement(47015. .47353)  
/rpt_family="ORR1C"  
repeat_region      47201. .47258  
/rpt_unit="(AG)x28  
repeat_region      47202. .47258  
/rpt_family="(GA)n"  
repeat_region      complement(47350. .47483)  
/rpt_family="RMER1B"  
repeat_region      complement(47643. .48403)  
/rpt_family="RMER1B"  
repeat_region      complement(48459. .48558)  
/rpt_family="ORR1C"  
repeat_region      48573. .48751  
/rpt_family="MERVL"  
repeat_region      48611. .48933  
/rpt_family="MERVL"  
repeat_region      48940. .49259  
/rpt_family="L1_MM"  
repeat_region      49288. .49549  
/rpt_family="MERVL"  
repeat_region      complement(49570. .49703)  
/rpt_family="(TA)n"  
repeat_region      49751. .49830  
/rpt_family="B1-F"  
repeat_region      complement(49884. .50083)  
/rpt_family="L1ME2"  
repeat_region      complement(49885. .50110)  
/rpt_family="L1ME1"  
repeat_region      51858. .51878  
/rpt_family="AT_rich"  
repeat_region      52211. .52307  
/rpt_family="PB1D9"  
repeat_region      52224. .52294  
/rpt_family="B1_MM"  
repeat_region      52319. .52422  
/rpt_family="L1ME2-Off2"  
repeat_region      52424. .52487  
/rpt_family="B1-F"
```

Query Match 12.6%; Score 132.2; DB 94; Length 72831;
Best Local Similarity 91.5%; Pred. No. 1.7e-27;
Matches 140; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY	588	gcctctcgtcaagaagatctggccaagatctcctcgtatattatcactggagac	647
Db	41738	gcttctcgtcattagaactcttgaaccaaagacattctcccaactgttgattcttggagct	41679
QY	648	ggcaagcttggatgaatctggaagcgcttattggtggagaactctgaacaatcagaagatgc	707
Db	41678	ggcagcgtttggatgaactctgggaacgcttattggaagaagacactctgaacacttgcagatgc	41619
QY	708	ttctaagatcctctgtctgacgctctgcacaagctt	740
Db	41618	ttcttaaacatctctgttgatctctctccaaaagct	41586

```

RESULT 10
AL512584/C
LOCUS      AL512584      162337 bp      DNA
DEFINITION Mus musculus chromosome 2 clone RP23-232D1, *** SEQUENCING IN
PROGRESS ***, In unordered pieces.
ACCESSION  AL512584
VERSION     AL512584.8  GI:13568243
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN.
SOURCE      house mouse.

```

REFERENCE
1 (pages 1 to 162337)
Clarke, D., Connor, R., leaves, N.I., Caveberry, L., Greystrom, J.,
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
Clarke, D., Connor, R., leaves, N.I., Caveberry, L., Greystrom, J.,

North, P.C., Hunter, G., Shuffiebottom, L., Kimberly, C., Campbell, M., Jones, S., Lawrence, N., Strachan, G.L., Greenham, L., Maggott, K. and Botcherby, M.R.M.

TITLE Direct Submission
JOURNAL Submitted (07-APR-2001) Mouse Sequencing Group, HGMPC-RC, Hinxton, Cambridge, CH10 1SB, UK. E-mail enquiries:- mtbotchehgmprc.ac.uk or dclakehgmprc.mrc.ac.uk
REMARK HGMPC-RC part of the UK Mouse Sequencing Consortium
COMMENT On Apr 9, 2001 this sequence version replaced g1:33509418.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES	Location/Qualifiers
source	1. .162337

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-232D1"
/clone_11b="RPCR-23"
misc_feature
1..17159

```

```
misc_feature      /note="assembly_fragment:1105"
                  17260. 27057
misc_feature      /note="assembly_fragment:1744"
                  27158. 44579
```

```
misc_feature    /note="assembly_fragment:1907"
                44680..67784
misc_feature    /note="assembly_fragment:2452"
                67885..70409
misc_feature    /note="assembly_fragment:2570"
                70410..70410
```

```
misc_feature
70510. 72696
/note="assembly_fragment:272"
72797. 112315
/note="assembly_fragment:3029"
```

```
misc_feature      112416 .114440 /note="assembly_fragment:322"  
misc_feature      114541 .131877 /note="assembly_fragment:3832"
```

```
misc_feature 131978. 134323 /note="assembly_fragment:590"
misc_feature 134424. 146222 /note="assembly_fragment:775"
misc_feature 146223. 150025 /note="assembly_fragment:776"
```

```
misc_creature 146323 /note="assembly_fragment:778"
BASE COUNT 43514 a 35098 c 35071 g 42212 t 6442 others
```

ORIGIN

Query Match	5.4%;	Score 57;	DB 81;	Length 162337;
Best Local Similarity	57.4%;	Pred. No. 3.1e-05;		
Matches 182; Conservative	0;	Mismatches 100;	Indels 35;	Gaps 3;

Qy 30 ggtgcgagtgcttlttgagggaacgagtttaglatatlaaacaccgggcgttgttllt 89
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21912 GGCAGCTGCTTTGAGGGAGCAATCCATCCTAGGAAGATCGGACGGGCTTCTGGCTT 21953

QY 90 ccgaatgtaattggcctataccagagagcgatccacgcggtgatgtcat 149
| | | | | | | | | | | | | | | | | | | | |
Db 21852 CTGAATGAAAGATTGA-CTGTGCTAGGTGTACAAATCCACACA--GCCACACTGGCCAG 21796

```

oy      150 gggtcttcctcgtgaagaagacccttcccttgccaggacatcggyglaaccgtttca    209
        ||| ||| ||||| ||||| ||| ||| ||| ||| ||| |
Db  21795 ggcgtttcccctgaaagaagaccttatcatggcctggacttgag-----   21750

```

Db 21751 -----GTAACTCATCTTACAGTGAAGAAAGACAGACGATTATTGCG 21708

Db 21707 ACACAGTGCATCTCTACTCTTGGAGATCGAGTTCCTTTAGTCTTGACAACTACTGC 21648

Db 21647 AGAGTCCACCCACCTCT 21631

RESULT	11
AC014256/c	
Locus	
Definition	AC014256 76748 bp DNA HTG 16-NOV-1999 Drosophila melanogaster; *** SEQUENCING IN PROGRESS ***, In ordered pieces.
Accession	AC014256
Version	AC014256.1 GI:6437079
Keywords	HTG; HTGS-PHASE2.
Source	Fruit fly.

ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephyridae; Drosophilidae; Drosophila.
1 (bases 1 to 76748)
AUTHORS
Adams, M. and Venter, J.C.
TITLE
Direct Submission
JOURNAL
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,

COMMENT
Rockville, MD, USA
This sequence was identified as CDW:10211789, by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers

```

SOURCE          1..70/46
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
BASE COUNT      21897 a 15665 c 15830 g 23356 t
ORIGIN

```

Query Match	5.48;	Score 56.2;	DB 63;	Length 76748;
Best Local Similarity	58.88;	Pred. No. 4.4e-05;		

[illegible]

924 catggtacttggatgaatgacgccttggccttggctgtgattatcacaccttacaatat 983
Db 1/052 CACACCAACCCCTACCAACCTGGCCGCTACACGACGACGACACCCGGTGCATCTCAA 16993


```

IMREYKHSYDRVSTREPTLYLEKKMLTFOILLALNOCHKOKICHGDKIKENTLITS
MNNILSDPASKPTIYIPEDNPADITYTPPDSRRRTCTYIAPREFKYTILASDDGNGN
MSVITHDSITIRLAPSYAGNTLIPAMDITPSAGALLIEMTGTAPPELSOLATVGRGR
DLVEKHLAIGENERLRLNLSMIDHSMNRSAEDYLDQERQGLPEEYFSLDSYD
MFSEPTIMSPDQIQCCLHNDICIKVLTQADIPPEEKEKQDPLCARLPEND
GLIITITVYSCIRGLKQSNWKIAMELLOKIKSYTTFETILDRILIPYLAKQSPA
KVOVOAIEVTLACIGAMKIDIPRSDANPEVYLTSTIDLASETSAYIVAFANISA
LAKSAYFLEETORANPNMDPTREYPAELNMLHDIIVRAQVSLTLDSDPVVKORLMS
GICDLCAPGKEKANDVILSHIMTFLINDEBNLNGAFYDNLGAVGYGMOASDILP
LLOOGTDEPEEVIKAKIRAVITLIEIGHIKPATEIYOEIRACIYCHGNALVEIC
GMATTRNLSDIDYCKIMPAIGAFKPLILOYEKHILLDCVHPPEPRPIEVSUR
FQDIHHTALEANSRVRSQTRQALPQTEEGQTLNLRFRUSBSGLJDLIEQJLA
MNPFLISMKHAKLQDLDTASNGRIYVSRQVACHEYPLADASKMPLIDRSEGET
LASPDADAITPLAGAGIAGSPASGAVLVGVTAVSAATSLGEYTMERNCMENERS
ECRKLDESLTAKISRYNVALISORCSYDKLVPTPEPMRMTGTALVLAHSESVIK
LASLRPHGSLFASGIDGYVRLMDCSKLNGONGKRSROYYSANPITYALACDSGGS
LAVGKDSGSMIMMRIDRNSAKMTLOALNONDHKGDVYMHAYDQATQSYVATLY
GTVIADPTMOHSAMRLONELRHGVTYTTTCADIPGSMLANCHSGCKHICOMLERLPI
AETKHPADSMIRKVAACHTEPSTILSASOSNNERYVNIETGQRQVIMASPVAAISS
ASPSDATTGCTLSGVVDSPTLTGSSDQIRYWDITNFKNSQIVKIPANDNDLVA
FNGAGLIDGSOVIEEQIISMASTGDSQDLRSTIEENRSGPDMPTASHDAIITDLML
CKRKGQIYIASASRNGYIKLWK"
  join(9619..10588,10978..11126,12073..12352,12809..13132,
13366..13591)
  /gene="CG8420"
  /product="CP24697"
  /db_xref="FLYBASE:FBan0008420"
  /db_xref="FLYBASE:FBgn0037664"
  <9619..>13591
  /gene="CG8420"
  /map="85D2-85D4"
  /db_xref="FLYBASE:FBan0008420"
  /db_xref="FLYBASE:FBgn0037664"
  join(9827..10588,10978..11126,12073..12352,12809..13132,
13366..13446)
  /note="CG8420 gene product"
  /codon_start=1
  /db_xref="FLYBASE:FBan0008420"
  /db_xref="FLYBASE:FBgn0037664"
  /protein_id="AAFS434.1"
  /db_xref="GI:7299145"
  /translation="MYRALLIACLLIPLGSGMSADESNPILDMASMPFOEALS
NONGNNGGAGAGLAVASLITFPQASKSGGAGAGGAGANMILISGLISLSS
QGQSGGSDPTIGVLEMTQGDDEATPOOKSKSGGSESGIGLITLLOVAFANM
TGQNDKATHHQKRTATEPESENGLMLPLWQAVSSFAPEQSOTQEKHSHAW
VLPPLEHIVLMDHFSNSELADALYERKGVKIMKPGKSDKLDYKLFESLNNGS
FRFRMTKATYTLADWASYLANPEVYLRFTQAIKNGKLSQGGPKQTHRDSRPG
ETISNLDIVAKHHLNWKIDSQYRPAVAGYKELKIQARQILKOROHNTESISKLD
TLNLEIEPVLRVHAYRYTSKPOCDRVYLCOLNMAALDQOKROHNDYVOOHOKP
QSOLENPTSSSLIAGVSPKIVIGSMGAIIFISTEGTPEFWLFCVINAIPNCEAVY
PVCNCFHSGEAKVTTEYIHNL"
  join(14006..14356,14503..>15118)
  /gene="CG16733"
  /product="CT37237"
  /db_xref="FLYBASE:FBan0016733"
  /db_xref="FLYBASE:FBgn0037665"
  <14006..>15118
  /gene="CG16733"
  /map="85D2-85D4"
  /db_xref="FLYBASE:FBan0016733"
  /db_xref="FLYBASE:FBgn0037665"
  join(14024..14358,14503..15118)
  /gene="CG16733"
  /note="CG16733 gene product"
  /codon_start=1
  /db_xref="FLYBASE:FBan0016733"
  /db_xref="FLYBASE:FBgn0037665"
  /protein_id="AAFS434.1"
  /db_xref="GI:7299146"
  /translation="MQLYRLEEDILRRTNAPVPONCEVLPDQFLIRKVEYLG
ESIRSLPVQDVWVSYPRGTSTMAQEWMLGHQDLYVAADQDLRLSPLEISLAL
FSIDHETVAQKFGNTVDLVNLRPRRPARSHLPWLLPQGFETVPRIVYATARNPD

```

```

LCVSYHYEKLHGMNGDEFOFVDLELGHTPMGSYRHVLPFKRSDONVLEIYE
DMWKDLPVVRRCARPLGYSOLDSYSTLOKLCQHLTEQDMRANKVNIENKILBESSK
FLRNGKIDMRNMGNMESERDEWTEHNRNGSLNPDV"
  join(<16163..16213,16307..16490,16637..>16836)
  /gene="CG16735"
  /product="CT37245"
  /db_xref="FLYBASE:FBan0016735"
  /db_xref="FLYBASE:FBgn0037666"
  <16163..>16836
  /gene="CG16735"
  /map="85D2-85D4"
  /db_xref="FLYBASE:FBan0016735"
  /db_xref="FLYBASE:FBgn0037666"
  join(16163..16213,16307..16490,16637..16836)
  /gene="CG16735"

```

```

Query Match 5.4% Score 56.2; DB 5; Length 224400;
Best local similarity 58.8%; Pred. No. 5.7e-05;
Matches 97; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 864 cccagcaagtcctataacctgcatataagtaatttgaatttcgtglttcaa 923
Db 2413 CACCAACCAATCCCTACACACTGCGCTCTACTACGCGACGACACCGGCTGATCTTCA 24192

QY 924 catgtacttggatgatgatgccttggccttggccttggctgtgattacacccctacaat 983
Db 24193 CATCATCTCTGTGCTTACGTGCTGTGACACTGTCTGTGCGGTATCTGTCACCCAT 24252

QY 984 ttgaacatgagcttgatgatgatacatatttaagatgac 1028
Db 24253 CGCTCATGATGATCCGCGACGAGACTCCATCATCATTTTACCCCATGAC 24297

```

```

RESULT 15
HSAC002069
LOCUS HSAC002069 158537 bp DNA PRI 09-MAY-1997
DEFINITION Human BAC clone RG326K09 from 7q21, complete sequence.
ACCESSION AC002069
VERSION AC002069.1 GI:2076718
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 158537)
AUTHORS Geisel,C., Beck,C., Smith,A and Twyman,B.
TITLE The sequence of H. sapiens BAC clone RG326K09
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 158537)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1997)
COMMENT Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA
http://genome.wustl.edu/gsc
e-mail: sapiens@watson.wustl.edu

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping

Project and the Washington University Genome Sequencing Center.
For additional information about the map position of this sequence,
see <http://www.ngri.nih.gov/DIR/GRB/CHR/> or send an E-mail to
egreen@ngri.nih.gov

SOURCE INFORMATION:

This clone is from a release of the human BAC library. The library
contains cloned DNA from a human sperm. For references see:
Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et
al., Genomics 34:213-218 (1996).

VECTOR: pBelo

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of H.RG326K09;
actual end is at 158537 of H.RG326K09. This clone lies in an
unanchored cluster of unknown orientation.

This clone contains SWS3233 (NID:q113712).

Location/Qualifiers

FEATURES

source

1. 158537

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/clone="RG326K09"

/clone_lib="CITB-HS-A"

/map="7q21"

276..1999

/rpt_family="L1"

complement(580..998)

/rpt_family="L1"

3260..3282

/rpt_family="L1"

4082..4178

/rpt_family="L1"

4193..4376

/rpt_family="L1"

4553..4725

/rpt_family="L1"

4767..4808

/rpt_family="L1"

5097..5621

/rpt_family="L1"

complement(5236..5642)

/rpt_family="L1"

6764..6910

/rpt_family="L1"

6922..6974

/rpt_family="L1"

complement(6997..7223)

/rpt_family="ALU"

7606..7712

/rpt_family="L1"

7866..7936

/rpt_family="L1"

11439..11461

/rpt_family="L1"

11641..11715

/rpt_family="L1"

12310..12397

/rpt_family="L1"

12589..12629

/rpt_family="L1"

13203..13236

/rpt_family="L1"

13854..14031

/rpt_family="L1"

16674..17204

/rpt_family="MER"

complement(17532..17751)

/rpt_family="L1"

complement(19040..19066)

/rpt_family="L1"

misc_feature

20125..21945

/note="Probable pseudogene. Similar to hsrp1alpha U28386

(NID:g899538)."

20132..20288

/note="Similar to human EST W56840 (NID:g1358716)

zc97d02.r1"

20278..20388

/note="Similar to human EST W56840 (NID:g1358716)

zc97d02.r1"

complement(26748..28317)

/rpt_family="THR"

29378..35406

/rpt_family="L1"

complement(33995..34413)

/rpt_family="L1"

complement(35884..35905)

/rpt_family="L1"

complement(35910..36194)

/rpt_family="ALU"

40798..45213

/note="Probable retroviral pseudogene. Similar to POL_GALV

POL POLYPROTEIN P21414 (NID:g130602)."

complement(46549..47052)

/rpt_family="MER"

complement(47102..47157)

/rpt_family="MER"

47521..50154

/rpt_family="L1"

51874..52163

/rpt_family="ALU"

53038..53063

/rpt_family="L1"

complement(56184..56476)

/rpt_family="ALU"

56478..56736

/rpt_family="ALU"

59292..60851

/rpt_family="THR"

61080..61206

/rpt_family="L1"

complement(65139..65325)

/rpt_family="L1"

75780..76108

/rpt_family="L1"

78275..78293

/rpt_family="L1"

79752..79781

/rpt_family="L1"

80350..80411

/rpt_family="ALU"

80412..80438

/rpt_family="L1"

80761..83307

/rpt_family="L1"

83308..83601

/rpt_family="ALU"

83602..83652

/rpt_family="L1"

83744..83767

/rpt_family="L1"

83784..83813

/rpt_family="L1"

84400..84452

/rpt_family="L1"

84599..84655

/rpt_family="L1"

complement(85751..85835)

/rpt_family="L1"

85770..85847

/rpt_family="L1"

89152..89193

/rpt_family="L1"

93896..93929

```
repeat_region      /rpl_family="L1"      complement(95512      .95708)
repeat_region      /rpl_family="ALU"      97302      97346
repeat_region      /rpl_family="L1"      98231      .98289
repeat_region      /rpl_family="L1"      98500      .98555
repeat_region      /rpl_family="L1"      complement(9131      .99423)
repeat_region      /rpl_family="ALU"      107834      107888
repeat_region      /rpl_family="L1"      108592      108680
repeat_region      /rpl_family="ALU"      108881      108923

Query Match      4.8%; Score 50.8; DB 93; Length 158537;
Best Local Similarity 56.3%; Pred. No. 0.002;
Matches 116; Conservative 0; Mismatches 87; Indels 3; Gaps 1;

OY 388 agtgaagaaagagtgtatgtaggaagaaactcagtggttgaagaccttcagtc 447
      ||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114499 AGTTAGCAATTAAACATATCTCAGCAATGATGTGACCAATGTATTCTCTCCAGCA 114558

OY 448 acctgcgcagcccgtaatgcctggttcaagaagaactcgtgttcagttcacctccc 507
      | ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 114559 AAGTTG---AAATTGCTAAATAAATTATCCCAAGGTAAAGACTCTCCTCACTTGTTGCTCT 114615

OY 508 ctcaattctcgaagtagaacaatgaagtgacctgctcttcttctgaactgcaagtg 567
      ||| ||| ||| ||| | | | | | | | | | | | | | | | | | | | | |
Db 114616 CTCTACTCTTTGAAGAATAACATATCCATCTCTTTCTTCTTGACACTTAAAAATC 114675

OY 568 ctacatgatattcaagcttgcctgc 593
      || ||||| ||||| ||||| |||
Db 114676 TCACCTGATATGTCACACTTGTGTC 114701
```

Search completed: August 18, 2001, 22:01:01
Job time: 13612 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: August 18, 2001, 19:15:13 ; Search time 250.24 Seconds
(without alignments)
2634.658 Million cell updates/sec

Title: US-09-284-320-31

Sequence: 1 atgctgcgtcttgcgtcgtc.....accagaatcgaatgcat 1050

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 730101 segs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_0601:*
2: /cgnl_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*
3: /cgnl_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*
4: /cgnl_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*
5: /cgnl_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*
6: /cgnl_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*
7: /cgnl_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*
8: /cgnl_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*
9: /cgnl_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*
10: /cgnl_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*
11: /cgnl_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*
12: /cgnl_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*
13: /cgnl_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*
14: /cgnl_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*
15: /cgnl_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*
16: /cgnl_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*
17: /cgnl_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*
18: /cgnl_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*
19: /cgnl_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*
20: /cgnl_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*
21: /cgnl_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*
22: /cgnl_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*
23: /cgnl_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1050	100.0	1050	19	AAV49561 Human epidermal ca
2	1050	100.0	2033	19	AAV49560 Human epidermal ca
3	1050	100.0	2108	22	AAV97893 Human secreted pro
4	1048.4	99.8	2092	22	AAV93774 Human cDNA encodin
5	1048.4	99.8	2099	22	AAV97929 Human secreted pro
6	971.8	92.6	2010	20	AAZ40848 Secreted protein E
7	390.8	37.2	560	22	AAV93970 Primer specific fo
8	386.2	36.8	858	20	AAZ15793 Human gene express
9	260.2	24.8	343	20	AAV51918 Human secreted pro
10	166.2	15.8	254	20	AAZ12587 Human gene express
11	153.8	14.6	469	21	AAV43203 Xenopus secreted e

C	12	83.2	7.9	936	22	AAF58252	Oligonucleotide D1
C	13	83.2	7.9	936	22	AAF58254	Oligonucleotide D1
C	14	83.2	7.9	936	22	AAF58257	Oligonucleotide D1
C	15	83.2	7.9	936	22	AAF58259	Oligonucleotide D2
C	16	83.2	7.9	936	22	AAF58262	Oligonucleotide D1
C	17	83.2	7.9	936	22	AAF58255	Oligonucleotide D1
C	18	76.2	7.3	936	22	AAF58252	Oligonucleotide D1
C	19	76.2	7.3	936	22	AAF58254	Oligonucleotide D1
C	20	76.2	7.3	936	22	AAF58257	Oligonucleotide D1
C	21	76.2	7.3	936	22	AAF58259	Oligonucleotide D2
C	22	76.2	7.3	936	22	AAF58262	Oligonucleotide D2
C	23	76.2	7.3	936	22	AAF58255	Oligonucleotide D1
C	24	46	4.4	418	21	AAV94060	Cat flea hindgut a
C	25	46	4.4	422	21	AAV94573	Cat flea hindgut a
C	26	44.4	4.2	379	21	AAV94468	Cat flea hindgut a
C	27	38.8	3.7	10342	19	AAV34684	Arabidopsis thalia
C	28	37	3.5	673	21	AAV37892	Arabidopsis thalia
C	29	37	3.5	676	21	AAV54354	Arabidopsis thalia
C	30	36.8	3.5	6297	21	AAV64395	DNA encoding a mur
C	31	36.6	3.5	18627	19	AAV52246	Streptococcus pneu
C	32	35.8	3.4	912	21	AAV48809	Arabidopsis thalia
C	33	35	3.3	4796	18	AAV67187	Huntingtin interac
C	34	35	3.3	4796	21	AAZ58746	Human huntingtin-1
C	35	34.6	3.3	3287	20	AAV13297	Enterococcus faeca
C	36	34.4	3.3	1242	18	AAV72169	Alzheimer's diseas
C	37	34.2	3.3	18359	20	AAV20255	Borrelia burgdorfe
C	38	34	3.2	389	21	AAV75518	Human ORF1073
C	39	34	3.2	1010	21	AAV34623	Arabidopsis thalia
C	40	33.8	3.2	244	22	AAV58238	Oligonucleotide D1
C	41	33.8	3.2	1181	21	AAV40766	Arabidopsis thalia
C	42	33.8	3.2	4525	21	AAV56975	Nucleotide sequenc
C	43	33.4	3.2	912	22	AAV5779	S. pneumoniae lacc
C	44	33.4	3.2	912	22	AAF29524	Streptococcus pneu
C	45	33.4	3.2	1934	19	AAV42999	Streptococcus pneu

ALIGNMENTS

RESULT 1
ID AAV49561 standard; cDNA to mRNA: 1050 BP.
AC AAV49561:
XX
XX
XX 21-OCT-1998 (first entry)
XX
XX
XX Human epidermal carcinoma cell line KB clone HP01293 cDNA #2.
DE
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentialiation; immune system; stimulator; suppressor; regulator;
KW haematopoiesis; activin; inhibitor; chemotactic; chemokine; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX
XX OS Homo sapiens.
XX
XX FN W09821328 f2.
XX
XX PD 22-MAY-1998.
XX
XX PF 07-NOV-1997; 97WO-JP04056.
XX
XX PR 13-NOV-1996; 96UP-0301429.
XX
XX PA (PROT-) PROTEGENE INC.
XX PA (SAGA) SAGAMI CHEM RES CENTRE.
XX
XX PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX WPI: 1998-297932/26.
XX DR P-PSDB: AAV64539.
XX
XX Human protein having transmembrane domain - useful for, e.g.


```

OY 121 cggatccagaagcgtgctgattgtccatgagcctctctgtgaagaagacctcttctg 180
DB 281 cggatccagaagcgtgctgattgtccatgagcctctctgtgaagaagacctctctg 340
OY 181 cccgagctcagcagtgaggtaacatgcttcatctgctcctgagctacacgcatatgtatgtg 240
DB 341 cccgagctcagcagtgaggtaacatgcttcatctgctcctgagctacacgcatatgtatgtg 400
OY 241 aagagagtgagaacaaactgctcctcagcagtgatcattcttacccttgagagat 300
DB 401 aagagagtgagaacaaactgctcctcagcagtgatcattcttacccttgagagat 460
OY 301 gcagcttcttctgcttctgacagtggtgcaaatcttcatcttacccttcttctgagaa 360
DB 461 gcagcttcttctgcttctgacagtggtgcaaatcttcatcttacccttcttctgagaa 520
OY 361 accctcgttcttctgcttctgacagtggtgcaaatcttcatcttacccttcttctgagaa 420
DB 521 accctcgttcttctgcttctgacagtggtgcaaatcttcatcttacccttcttctgagaa 580
OY 421 aactcagtgcttctgacagtggtgcaaatcttcatcttacccttcttctgagaa 480
DB 581 aactcagtgcttctgacagtggtgcaaatcttcatcttacccttcttctgagaa 640
OY 481 gaaactcgttctgacagtggtgcaaatcttcatcttacccttcttctgagaa 540
DB 641 gaaactcgttctgacagtggtgcaaatcttcatcttacccttcttctgagaa 700
OY 541 ctgctcttcttctgacagtggtgcaaatcttcatcttacccttcttctgagaa 600
DB 701 ctgctcttcttctgacagtggtgcaaatcttcatcttacccttcttctgagaa 760
OY 601 aagcatctcagcagtggtgcaaatcttcatcttacccttcttctgagaa 660
DB 761 aagcatctcagcagtggtgcaaatcttcatcttacccttcttctgagaa 820
OY 661 gaaatcggagcgttctgagagagacgtcgaacatctcagagatgcttctcgaagatcctt 720
DB 821 gaaatcggagcgttctgagagagacgtcgaacatctcagagatgcttctcgaagatcctt 880
OY 721 gttgagcgtcctcagagtggtgcaaatcttcatcttacccttcttctgagaa 780
DB 881 gttgagcgtcctcagagtggtgcaaatcttcatcttacccttcttctgagaa 940
OY 781 gtagagtgtagctcgtcgaatcttctgacacccctcctcctcctcctcctcctcctcctcct 840
DB 941 gtagagtgtagctcgtcgaatcttctgacacccctcctcctcctcctcctcctcctcctcct 1000
OY 841 ctgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 900
DB 1001 ctgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1060
OY 901 ttgtaaatctcgtgcttctcgaacatgcttctgaaatgctcctcctcctcctcctcctcctcct 960
DB 1061 ttgtaaatctcgtgcttctcgaacatgcttctgaaatgctcctcctcctcctcctcctcctcct 1120
OY 961 gttattatctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1020
DB 1121 gttattatctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1180
OY 1021 aggatgtacaacacagagatcctgaatgat 1050
DB 1181 aggatgtacaacacagagatcctgaatgat 1210

```

RESULT 5

AA#97929 standard; CDNA; 2099 BP.

AA#97929;

01-JUN-2001 (first entry)

```

XX Human secreted protein CDNA, SEQ ID NO: 56.
XX
XX Human: secreted protein; immunomodulatory; antisclerotic;
KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KW nototropic; anticonvulsant; antialzheimer's; antiparkinsonian;
KW antimicrobial; vulnery; vaccine; gene therapy; cancer;
KW protein coordinate data; infection; ss.
XX
OS Homo sapiens.
XX
XX WO200121658-A1.
XX
XX 29-MAR-2001.
XX
XX 22-SEP-2000; 2000WO-US26013.
XX
XX 24-SEP-1999; 99US-0155709.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ni J., Baker KP., Birse CE., Ebner R., Fiscella M., Komatsoulis GA.;
PI Tafleur DW., Moore PA., Olsen HS., Rosen CA., Ruben SA., Soppet DR.;
PI Young PE., Wei P., Florence KA.
XX
XX WPI: 2001-23511/24.
XX
XX Nucleic acids encoding 32 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy -
XX
XX Claim 1: Page 749-750; 890pp; English.
XX
XX The present sequence encodes one of 32 novel human secreted polypeptides.
XX The nucleic acid molecules and polypeptides they encode may be used in
XX the prevention, diagnosis and treatment of diseases such as
XX immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
XX and human immunodeficiency virus (HIV) infections), hyperproliferative
XX disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
XX (e.g. Scimitar syndrome, Chagas' cardiomyopathy and coronary
XX arteriosclerosis), angiogenic disorders (e.g. corneal graft
XX neovascularisation and diabetic retinopathy), neurological disorders
XX (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
XX infectious diseases and/or for promoting wound healing, regeneration
XX and/or chemotaxis. The nucleic acid molecules may be used to produce the
XX secreted polypeptides. They may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar nucleic acid
XX sequences in samples. The polypeptides may be used as antigens in the
XX production of antibodies and in assays to identify modulators of
XX their expression and activity.
XX
XX Sequence 2099 BP; 603 A; 373 C; 461 G; 656 T; 6 other:
XX
XX
XX Query Match 99.8%; Score 1048.4; DB 22; Length 2099;
XX Best Local Similarity 99.6%; Pred. No. 0;
XX Matches 1046; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
XX
XX OY 1 atgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 60
XX 114 atgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 173
XX 61 atataaatccagagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
XX 174 atataaatccagagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 233
XX OY 121 cggatccagaagcgtgctgattgtccatgagcctctctgtgaagaagaccttctctg 180
XX 234 cggatccagaagcgtgctgattgtccatgagcctctctgtgaagaagaccttctctg 293
XX OY 181 cggatccagaagcgtgctgattgtccatgagcctctctgtgaagaagaccttctctg 240

```

Db	294	ccaggacgcgagttgggtataacctgtttcaatcgtcctctgggtacacgtatgtatgtgtg	353
Oy	241	aaggagatgaaacaactggtcctataccccagagcaatgtacatttcgtaacctttgagat	300
Db	354	aaggagatgaaacaactggtcctataccccagagcaatgtacatttcgtaacctttgagat	413
Oy	301	gaagttcccttttagcttgaacagatgttgaaatctcaatctcaactctatctttcgaaga	360
Db	414	gaagttcccttttagcttgaacagatgttgaaatctcaatctcaactctatctttcgaaga	473
Oy	361	actcctgtgttttgcaagtgtgctcccaagtgaagaagatgtatatatgttaaggaga	420
Db	474	actcctgtgttttgcaagtgtgctcccaagtgaagaagatgtatatatgttaaggaga	533
Oy	421	aactcaagtttgaagaaccttttcagtaacctttgacgaagctccgaatccgtgttttcaa	480
Db	534	aactcaagtttgaagaaccttttcagtaacctttgacgaagctccgaatccgtgttttcaa	593
Oy	481	gaaaacctgtgttctcaagtttaacctccccctaattctctcgaagtagaacaagaagtac	540
Db	594	gaaaacctgtgttctcaagtttaacctccccctaattctctcgaagtagaacaagaagtac	653
Oy	541	cgcgccttcttcttgaactgaactgaagtgtacatgtatcttaagaactgtgtctgcgcac	600
Db	654	cgcgccttcttcttgaactgaactgaagtgtacatgtatcttaagaactgtgtctgcgcac	713
Oy	601	aagcactcagcagaagatcaattccctcgtattatattactcgaagctctggcagatttgat	660
Db	714	aagcactcagcagaagatcaattccctcgtattatattactcgaagctctggcagatttgat	773
Oy	721	gttcgacgtctcgaaaaagtttgcagatgaatgaatgaatcagtttatgttggaaagcag	780
Db	834	gttcgacgtctcgaaaaagtttgcagatgaatgaatgaatcagtttatgttggaaagcag	893
Oy	781	gtagaagttagtcacgtcgaagtaatttgaaacctccctcaattagaagaacaagatactc	840
Db	894	gtagaagttagtcacgtcgaagtaatttgaaacctccctcaattagaagaacaagatactc	953
Oy	841	cttggaggaacaacaagcgaagaacccagcaagctccctataaccttcataaagatnaat	900
Db	954	cttggaggaacaacaagcgaagaacccagcaagctccctataaccttcataaagatnaat	1011
Oy	901	tttgaataatccgttggttttcaacaagtatactttgataatgataatgccttggccttgct	960
Db	1014	tttgaataatccgttggttttcaacaagtatactttgataatgataatgccttggccttgct	1077
Oy	961	ggtatataaccccttcaacaattttgnaaacaagatccttgatataatgaatgaatcaat	1020
Db	1074	ggtatataaccccttcaacaattttgnaaacaagatccttgatataatgaatgaatcaat	1133
Oy	1021	aggatgacaacaacagaagattcgaaatgat	1050
Db	1134	aggatgacaacaacagaagattcgaaatgat	1163

RESULT	6
AAZ40848	
ID	AAZ40848 standard; DNA; 2010 BP.
XX	
AC	
XX	AAZ40848;
XX	
DT	18-JAN-2000 (first entry)
DE	Secreted protein EST coding sequence 33-774-E2-FL1.
XX	
XX	
KM	Secreted protein; fingerprint identification technique;
KM	chromosome mapping; human; hereditary disease; diagnosis; cancer;
KM	hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
KM	autoimmune disease; rheumatic disease; embryonic disorder; myopathy

KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
 KW hypertension; ss.
 OS Homo sapiens.
 PN M09940189-A2.
 XX
 XX
 PD 12-AUG-1999.
 XX
 PF 09-FEB-1999; 99MO-IB00282.
 XX
 PR 09-FEB-1998; 98US-0074121.
 PR 13-APR-1998; 98US-0081553.
 PR 10-AUG-1998; 98US-0096116.
 PR 04-SEP-1998; 98US-0099273.
 XX
 PA (GESP) GENSET.
 PL Bougueleret L, Duclert A, Dumas Milne Edwards J;
 XX
 DR WPI; 1999-600966/51.
 DR P-PSDB; AAY59720.
 XX
 PT Extended cDNAs useful for expressing secreted proteins and to obtain
 specific antibodies -
 XX
 PS
 PS
 XX Claim 1: Page 227-229; 244pp: English.
 CC This sequence encodes a human secreted protein of the invention. The
 CC extended cDNAs (or genomic DNAs obtainable from them) may be used to
 CC prepare PCR primers and probes. These are useful for forensic matching or
 CC positive identification by DNA sequencing. They may also be used in
 CC alternative fingerprint identification techniques. Antibodies against the
 CC proteins encoded by the extended cDNAs are useful in identification of
 CC tissue types or cell species, as well as identifying tissue specific
 CC soluble proteins. The sequences can be used for chromosome mapping and
 CC identification of genes associated with hereditary diseases or drug
 CC response. Signal sequences from the cDNAs can be used in construction of
 CC secretion vectors. Other sequences derived from the extended cDNAs can be
 CC used to clone upstream genomic DNA sequences including promoters. This is
 CC in turn useful for identifying proteins that interact with promoter
 CC sequences. Some of the proteins may be useful in diagnosing and treating
 CC several disorders including, but not limited to: cancer, hyperlipidaemia,
 CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
 CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
 CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.
 XX
 SQ Sequence 2010 BP; 565 A; 360 C; 431 G; 642 T; 12 other;

Query Match	Similarity	92.64;	Score	971.8;	DB	20;	Length	2010;
Best Local	Similarity	96.98;	Pred.	No.	3.5e-296;			
Matches	1015;	Conservative	6;	Mismatches	23;	Indels	3;	Gaps
QY	1 atgagctgctgtctgcgtctcctccgcgacgtatggtgagagacgaagttcag							
Db	65 atgagcgtgtctgtcgtcgtccgcgcggtatgttgccggtcttgaggaagttcag							
QY	61 atattaaatccacccaggtctgctgttcttcgcgaatggaattgctctataccaggaag							
Db	125 atattaaatccacccaggggctcgtctgtcttcgcgaaggaattgctctataccaggaag							
QY	121 cggatccacgaacgtgctgcattcgtccatlgagagctctctctgtgaagaagaccttctcgg							
Db	185 cggatccacgaacgtgctgcattcgtccatlgagagctctctctgtgaagaagaccttctcgg							
QY	181 ccagagactcgcagttgagtaacctgttatctgctccgcgcacacgcgtcatgtagtggtg							
Db	245 ccagagactcgcagttgagtaacctgtttatctgctccgcgcacaggtccatgtagtagtggtg							
QY	241 aaggagatggaacaaactgctctatcccccagagcagtgtcattctgcaccttggagaat							
Db	305 aaggagatggaacaaactcctctatcccccagagcagtgtcattctgcaccttggagaat							

Oy	301	gatttcccttttagctgacgcaatgtgtgcaatttcatactctctattttctgaagaa	360
Db	365	gcattctctttagtcttgacagtggtgtgcaatttcatactctctattttctgaagaa	424
Oy	361	actccgtgttttgcagttgagtcgtccagtgagaaagatgtatagtlaaggaagca	420
Db	425	actccgtgttttgcagttgagtcgtccagtgagaaagatgtatagtlaaggaagca	484
Oy	421	aaccagtggtttgaagacatttcagtaacctctggccagctccgtctaaagccgtgttca	480
Db	485	aaccagtg-tggaaracctttagtaacctctggccagctccgtctaaagccgtgttca	543
Oy	481	gaaacctggtctcagttcaatcccccctaattc-tctagtagaacaatgaagtgtga	539
Db	544	gaaacctggtctcagttcaatcccccctaattcagtaagtagaacaatgaagtgtga	603
Oy	540	cctctcttcttcttcagtaactgcaatgtctcaatgatatctaaagctgtgtctgcgca	599
Db	604	ccgctcttcttcttcagtaactgcaatgtctcaatgatatctaaagctgtgtctgcgca	663
Oy	600	taagcatctagcccaagatcatcttcccgatttatatctacgtgagctgtgcagttgtga	659
Db	664	taagcatctagcccaagatcatcttcccgatttatatctacgtgagctgtgcagttgtga	723
Oy	660	tgaatttggaagcgtaatgtggaggaagactctgcaacaattcagaagatgctctaaagatcct	719
Db	724	tgaatttggaagcgtaatgtggaggaagactctgcaacaattcagaagatgctctaaagatcct	783
Oy	720	tgttgagcgctctgcgcaaaagtttgcagatgagacatgtacgtcttatgttgaggaatgcaggt	779
Db	784	tgttgagcgctctgcgcaaaagtttgcagatgagacatgtacgtcttatgttgaggaatgcaggt	843
Oy	780	ggttagagttagtaactgtgtcaagtcatttttgacaacctccctcaatttaggaagaacaagactat	839
Db	844	ggttagagttagtaactgtgtcaagtcatttttgacaacctccctcaatttaggaagaacaagactat	903
Oy	840	ccttgagcgcaaaacaagaggaagaccgcagccagccctacatcccttcagatataagataa	899
Db	904	ccttgagcgcaaaacaagaggaagaccgcagccagccctacatcccttcagatataagataa	963
Oy	900	ttttggaattatccgtgtgttttcaaatgtgtaacttttgataaagatgcctgtgccttgc	959
Db	964	ttttggaattatccgtgtgttttcaaatgtgtaacttttgataaagatgcctgtgccttgc	1023
Oy	960	tgtgattatcacctcttcaataatatttggaagaatggt-attcctgatatgatatgatacttt	1018
Db	1024	tgtgattatcacctcttcaataatatttggaagaatggaatccttgatatgatatgatacttt	1083
Oy	1019	atagagatgacaacacagagaattcgga	1045
Db	1084	atagagatgacaacacagagaattcgga	1110
RESUL	7		
ID	AAF93970		
AC	AAF93970 standard; DNA; 560 BP.		
XX	AAF93970;		
XX	23-MAY-2001 (first entry)		
DT			
XX			
DE			
XX			
KW	Human; secretory protein; membrane protein; vacuole; gene therapy;		
XX	rheumatoid arthritis; diabetes; PCR primer; ss.		
OS	Synthetic.		
XX			
PN	EP1067182-A2.		
XX			
PD	10-JAN-2001.		
XX			

PF	07-JUL-2000:	200DEP-0114090.
XX	PR	08-JUL-1999: 99JP-0194179.
XX	PR	11-JAN-2000: 2000JP-0118775.
XX	PR	02-MAY-2000: 2000JP-0183766.
PA	(HELI-) HELIX RES INST.	
XX	Oka T., Isogai T., Nishikawa T., Kawai Y., Sugiyama T., Hayashi K.	
PI	WPI: 2001-093989/11.	
DR	Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development -	
XX	Claim 4: SEQ ID 404; 609pp + CD ROM; English.	
PS	CC This invention relates to nucleic acid sequences AAF93744 - AAF93916	
XX	CC which encode human secretory or membrane proteins represented by	
CC	CC AAB88317 - AAB88419. Included in the invention are primers	
CC	CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the	
CC	CC cDNA sequences of the invention. The invention also includes methods for	
CC	CC the production of antibodies directed against the proteins, and cDNA	
CC	CC sequences, which can be used in vaccines. The polynucleotide sequences	
CC	CC can be used in gene therapy. The polynucleotide sequences and the	
CC	CC proteins they encode may be used in the prevention, treatment and	
CC	CC diagnosis of diseases associated with inappropriate secretory	
CC	CC protein/membrane protein expression. The nucleic acids and complementary	
CC	CC sequences may also be used as DNA probes in diagnostic assays	
CC	CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the	
CC	CC presence of similar nucleic acid sequences in samples. They may also be	
CC	CC used to study the expression and function of secretory proteins/membrane	
CC	CC polypeptides and their role in metabolism. The polypeptides may be used	
CC	CC as antigens in the production of antibodies against them and in assays to	
CC	CC identify modulators (agonists and antagonists) of expression and	
CC	CC activity. The antibodies and antagonists may also be used as therapeutic	
CC	CC agents to down regulate expression and activity. The antibodies may also	
CC	CC be used as diagnostic agents for detecting the presence of the	
CC	CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay	
CC	CC (ELISA). Examples of diseases which may be treated include rheumatoid	
CC	CC arthritis and diabetes.	
XX	Sequence 560 BP; 104 A; 145 C; 163 G; 142 T; 6 other;	
SO		
Query Match:	37.28; Score 390.8; DB 22; Length 560;	
Best Local Similarity	98.28; Pred. No. 5,6e-114;	
Matches 392: Conservative	0; Mismatches 7; Indels 0; Gaps 0;	
OY	1 atgagctggttttcgtcgtccctcgacggttggtagtgatgattgcgaagaagttaagt 60	
DB	162 atgagctggttttcgtcgtccctcgacggttggtagtgatgattgcgaagaagttaagt 221	
OY	61 atataaatcacccagggtctgtgttttcogaatggaaatgtgacctataccaggagag 120	
DB	222 atataaatcacccagggtctgtgttttcogaatggaaatgtgacctataccaggagag 281	
OY	121 cgagaccacaagcttgctgcacatgcatcgaaggctcttcctgtaagaagaagacttctcgg 180	
DB	282 cgagaccacaagcttgctgcacatgcatcgaaggctcttcctgtaagaagaagacttctcgg 341	
OY	181 ccagagactcgaagtgtgftaacacctgttatcgtctcctcggcaccacgttacatgtagtagtg 240	
DB	342 ccagagactcgaagtgtgftaacacctgttatcgtctcctcggcaccacgttacatgtagtagtg 401	
OY	241 aagggagatggaacaactggtctcaccgcccaaggacagttaacttctgtaaccttggagaat 300	
DB	402 aagggagatggaacaactggtctcaccgcccaaggacagttaacttctgtaaccttggagaat 461	
OY	301 ggaagctccttttgtcttgaaagagtttgaatatccaactcaactcctaattttttaaggaa 360	
DB	462 ggaagctccttttgtcttgaaagagtttgaatatccaactcaactcctaattttttaaggaa 521	

QY 361 acctcgttttttgcagttgctcccaaggagaaga 399
 |||||
 DB 522 acctcgttttttgcagttgctcccaaggagaaga 560

RESULT 8

AAZ15793 standard; cDNA; 858 BP.

AAZ15793;

12-OCT-1999 (first entry)

Human gene expression product cDNA sequence SEQ ID NO:3262.

Human: gene: gene expression product; diagnosis; therapy; probe; detection; mapping; tissue typing; profiling; forensic; cancer; genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

Homo sapiens.

MO9938972-A2.

05-AUG-1999.

28-JAN-1999: 99MO-US01619.

03-APR-1998: 98US-0080666.

28-JAN-1998: 98US-0072910.

24-FEB-1998: 98US-0075954.

31-MAR-1998: 98US-0080114.

03-APR-1998: 98US-0080515.

(CHIR) CHIRON CORP.

(HYSE-) HYSEQ INC.

Chkvenjkor R, Dickson M, Dmanac R, Dmanac S;

Becobedo J, Garcia PD, Garcia V, Giese K, Innis MA;

Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

Stache-Grain B, Sudduth-Klinger J, Williams LT;

WPI: 1999-494092/41.

Novel human genes and their expression products which are

differentially expressed in different cell types

Claim 1, Page 1565; 2479pp; English.

The present invention describes a library of human polynucleotides comprising the sequences given in AAZ12532 to AAZ17779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

Sequence 858 BP; 205 A; 186 C; 189 G; 225 T; 53 other;

Query Match 36.8%; Score 386.2; DB 20; Length 858;
 Best Local Similarity 92.4%; Pred. No. 2e-112;
 Matches 465; Conservative 0; Mismatches 30; Indels 8; Gaps 6;

QY 93 aaatgaattggtctataccagagagagatcccaagctgctgctatgctcag 152
 ||| |
 DB 59 aaancgttgggtcctataccagagagagatcccaagctgctgctatgctcag 118
 QY 153 ctctctgtgaagaagaaccttcttgccagagctcagagtggtggttaacctgttcatag 212
 |||||
 DB 119 ctctctgtgaagaagaaccttcttgccagagctcagagtggtggttaacctgttcatag 178
 QY 213 tccctgggtacagctcat-ggtgacgtgagagagtgagcaacaactggtctaccccccag 271
 |||||
 DB 179 ncncgggtacagctcatggtggtgagtgagtgagtgagtgagtgagtgagtgagtgag 238
 QY 272 gcaagtgatcttgctacagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 331
 |||||
 DB 239 gcaagtgatcttgctacagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 298
 QY 332 attcaatcactccttatttcttgaggaactcctgtgttggagtggtggtcctccagtg 391
 |||||
 DB 299 attcaatcactccttatttcttgaggaactcctgtgttggagtggtggtcctccagtg 358
 QY 392 aggaagaagtgatagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 451
 |||||
 DB 359 aggaagaagtgatagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 418
 QY 452 t--ggtccagctcgtgaatcgctgttccaagaactctgttccag-ttcaactccc 507
 |||||
 DB 419 ttgcgcgaagctcccgtaatcgctgttccaagaactctgttccaagaactcctcc 478
 QY 508 ctcaattctcgtgagtgaggaactg-aagttgacccgtc-ttcttctggaactcag 565
 |||||
 DB 479 ctcaattctcgtgagtgaggaactg-aagttgacccgtc-ttcttctggaactcag 538
 QY 566 tggctac-atgatatcttcaagctt 587
 |||||
 DB 539 tggctacatgatatttcaagctt 561

RESULT 9

AAZ151918 standard; DNA; 343 BP.

AAZ151918;

22-JUN-1999 (first entry)

Human secreted protein 5' EST SEQ ID NO: 132.

Human: secreted protein; EST; expressed sequence tag; diagnosis;

forensic; gene therapy; chromosome mapping; signal peptide;

upstream regulatory sequence; cytokine activity; cell proliferation;

differentiation; haematopoiesis regulation; tissue growth regulation;

reproductive hormone regulation; chemotaxis; chemokinetic; haemostatic;

thrombolytic; anti-inflammatory; tumour inhibition; ds.

Homo sapiens.

MO9906552-A2.

11-FEB-1999.

31-JUL-1998: 98MO-IB01236.

01-AUG-1997: 97US-0905223.

(GEST) GENSET.

Duclet A, Dumas Milne Edwards J, Lacroix B;

DR WPI: 1999-153782/13.
 DR P-PSDB; AA131118.
 XX
 PT New isolated brain-derived nucleic acids - used to develop products
 PT which may have cytokine, immune, regulatory, haematopoiesis
 PT regulating, anti-inflammatory or tumour inhibition activity
 PS Claim 1: Page 273-274; 577pp; English.
 XX
 CC AA151787 to AA152019, represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins, and encode the proteins given in AA112987 to
 CC AA13219, respectively. The proteins given represent the signal peptide
 CC and an N-terminal fragment of a secreted protein. The nucleic acid
 CC sequences can be used for producing secreted human gene products. They
 CC can also be used to develop products for diagnosis and therapy. The
 CC proteins obtained may have cytokine activity, cell
 CC proliferation/differentiation activity, haematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity or other activities. The products
 CC can be used in forensic, gene therapy and chromosome mapping procedures.
 CC The sequences can also be used for obtaining corresponding promoter
 CC sequences. The nucleic acids encoding the signal peptide can be used for
 CC directing extracellular secretion of a polypeptide or the insertion of a
 CC polypeptide into a membrane, or importing a polypeptide into a cell.
 XX
 SQ Sequence 343 BP; 56 A; 88 C; 104 G; 88 T; 7 other:
 XX
 Query Match 24.8%; Score 260.2; DB 20; Length 343;
 Best Local Similarity 96.4%; Pred. No. 1.3e-72;
 Matches 265; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 OY 1 atgctctgtgttgcgtgctcctggtggtggtggtgttggggagacagttagt 60
 DB 65 atgctctgtgttgcgtgctcctggtggtggtggtgttggggagacagttagt 124
 OY 61 atattaataatcaccagaggtctgttttccgaagaatggtccataccagaagag 120
 DB 125 atattaataatcaccagaggtctgttttccgaagaatggtccataccagaagag 184
 OY 121 cggatccagagcgtgctgcatgttcacatgggtctctcgttgaagaagaccttcttg 180
 DB 185 cggatccagagcgtgctgcatgttcacatgggtctctcgttgaagaagaccttcttg 244
 OY 181 ccaggaactcgagtggttaacctgttcatcgttcctcgagctaacgtatgtagtgcg 240
 DB 245 ccaggaactcgagtggttaacctgttcatcgttcctcgagctaacgtatgtagtgcg 304
 OY 241 aaggagatgaacaactgcttaccctccagagcag 275
 DB 305 aaggagatgaacaacttaccctccagagcag 339
 RESULT 10
 AA121587
 ID AA121587 standard; cDNA; 254 BP.
 XX
 AC AA121587;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:56.
 XX
 KW Human: gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 MO938972-A2.

PD 05-AUG-1999.
 XX
 PE 28-JAN-1999; 99MO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Chkrenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Glese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Grain B, Sudduth-Klinger J, Williams LT;
 XX
 DR WPI: 1999-494092/41.
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 XX
 PS Claim 1: Page 674; 2479pp; English.
 XX
 CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA121532 to AA17779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA121532 to AA17779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnosis (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 XX
 SQ Sequence 254 BP; 39 A; 50 C; 55 G; 63 T; 47 other:
 XX
 Query Match 15.8%; Score 166.2; DB 20; Length 254;
 Best Local Similarity 84.8%; Pred. NO. 7.7e-43;
 Matches 168; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
 OY 97 ggaattggcctataccagagagcgatcccaagcgtgctgcatgttcatgggttc 156
 DB 1 ggaattggcctataccagagagcgatcccaagcgtgctgcatgttcatgggttc 60
 OY 157 tcttgaagaagaaccttcttgccagagactcgagtggttaacctgttcatgctct 216
 DB 61 tcttgaagaagaaccttcttgccagagactcgagtggttaacctgttcatgctct 120
 OY 217 cgggtaccgcatgagtgatgagtggaagtggaacaactgctctaccacccagcagt 276
 DB 121 cgggtaccgcatgagtgatgagtggaagtggaacaactgctctaccacccagcagt 180
 OY 277 gtcattcgtacccttg 294
 DB 181 nmtltnatnccmng 198
 RESULT 11

AA	AAA43203	
ID	AAA43203	standard; cDNA; 469 BP.
XX		
AC	AAA43203;	
DT	21-MUG-2000	(first entry)
XX		
DE	Xenopus	secreted expressed sequence tag SEQ ID NO:1943.
KM	Human;	mouse; xenopus; rat; secreted expressed sequence tag; SESTF
KM	expressed	sequence tag; EST; probe; chemotactic; proliferative;
KM	immunomodulatory;	haematopoietic; chemokine; analgesic; haemostatic
KM	thrombolytic; antiinflammatory;	cytostatic; antibacterial; antifungal;
KM	antiviral; antidiabetic;	antisthmatic; vulnary; antiparkinsonian;
KM	antitumor; osteopathic;	neuroprotective; neurotropic; antiposrotatic;
KM	cardioprotective;	anticonvulsant; antidepressant; gene therapy;
KM	vaccine; autoimmune	disorder; multiple sclerosis; allergic condition;
KM	insulin dependent	diabetes; asthma; myeloid cell deficiency; ulcer;
KM	lymphoid cell	deficiency; burn; osteoporosis; osteoarthritis;
KM	central nervous	system disorder; Alzheimer's disease; stroke;
KM	Parinson's	disease; Huntington's disease; coagulation disorder;
KM	haemophilia;	thrombosis; inflammatory disorder; Crohn's disease;
KM	tumor;	infection; depression; psoriasis; SS.
XX		
OS	Xenopus	sp.
XX		
PN	W0200021990-A1.	
XX		
PD	20-APR-2000.	
XX		
PF	15-OCT-1999;	99WO-US24205.
XX		
PR	15-OCT-1998;	98US-0104435.
PA		
XX		
PA	(GENY) GENETICS INST INC.	
XX		
PI	Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;	
P1	Merberg D, Treacy M;	
XX		
DR	WPI; 2000-317937/27.	
XX		
PT	Isolated	polynucleotides, and encoded proteins, comprising secreted
PT	expressed	sequence tags (SESTs), useful for treating various disorders
PT	such as	autoimmune, infectious, and central nervous system disorders -
XX		
PS	Claim 1;	Page 564; 618PP; English.
XX		
CC	AAA41261	to AAA43419 represent specifically claimed secreted expressed
CC	sequence	tags (SESTs), isolated from human, mouse, xenopus and rat
CC	tissue	sources. The SESTs can have a range of activities depending on
CC	the	tissues they were isolated from. The activities include:
CC	chemotactic;	proliferative; immunomodulatory; haematopoietic;
CC	chemokineic;	analgesic; haemostatic; thrombolytic; antiinflammatory;
CC	cytostatic;	antibacterial; antifungal; antiviral; antidiabetic;
CC	antisthmatic;	vulnary; antitumor; osteopathic; neuroprotective;
CC	neurotropic;	antiparkinsonian; antiposrotatic; cardioprotective;
CC	anticonvulsant;	and antidepressant. The SESTs can be used for gene
CC	therapy	and in vaccines. The SESTs are useful as probes for the
CC	identification	and isolation of full-length cDNAs and genomic DNA
CC	molecules	which correspond to the SESTs. Proteins encoded by the SESTs
CC	are	useful in assays for determining biological activity and raising
CC	antibodies.	They may be useful for determining biological activity and raising
CC	(multiple	sclerosis; insulin dependent diabetes), allergic conditions
CC	(asthma),	myeloid or lymphoid cell deficiencies, wounds, burns, ulcers
CC	osteoporosis,	osteoarthritis, central nervous system disorders
CC	(Alzheimer's,	Parkinson's, Huntington's disease, stroke), coagulation
CC	disorders	(haemophilia, thrombosis), inflammatory disorders (Crohn's
CC	diseases),	tumors, bacterial, fungal or viral infections, depression and
CC	psoriasis.	AAA43420 to AAA43425 represent linker variants which are given
CC	in	the exemplification of the present invention.
XX		
XX	Sequence	469 BP; 107 A; 114 C; 109 G; 139 T; 0 other;
XX		

[illegible]

